Graded BDI Models for Agent Architectures *

Ana Casali[1], Lluís Godo[2] and Carles Sierra[2]

[1] Depto. de Sistemas e Informática
Facultad de Cs. Exactas, Ingeniería y Agrimensura
Universidad Nacional de Rosario
Av Pellegrini 250, 2000 Rosario, Argentina.

[2] Institut d'Investigació en Intel·ligència Artificial (IIIA) - CSIC
Campus Universitat Autònoma de Barcelona s/n
08193 Bellaterra, Catalunya, España.

Abstract. In the recent past, an increasing number of multiagent systems (MAS) have been designed and implemented to engineer complex distributed systems. Several previous works have proposed theories and architectures to give these systems a formal support. Among them, one of the most widely used is the BDI agent architecture presented by Rao and Georgeff. We consider that in order to apply agents in real domains, it is important for the formal models to incorporate a model to represent and reason under uncertainty. With that aim we introduce in this paper a general model for graded BDI agents, and an architecture, based on multi-context systems, able to model these graded mental attitudes. This architecture serves as a blueprint to design different kinds of particular agents. We illustrate the design process by formalising a simple travel assistant agent.

1 Introduction

In the recent past, an increasing number of multiagent systems (MAS) have been designed and implemented to engineer complex distributed systems. Several previous works have proposed theories and architectures to give these systems a formal support. Agent theories are essentially specifications of agents' behaviour expressed as the properties that agents should have. A formal representation of the properties helps the designer to reason about the expected behaviour of the system [25]. Agent architectures represent a middle point between specification and implementation. They identify the main functions that ultimately determine the agent's behaviour and define the interdependencies that exist among them [25]. Agent theories based on an intentional stance are among the most common ones. Intentional systems describe entities whose behaviour can be predicted

* A preliminary version of this paper, "Modelos BDI graduados para Arquitecturas de Agentes" (in Spanish), was presented at the Argentine Symposium on Artificial Intelligence (ASAI'04) and will appear in an especial issue of "Inteligencia Artificial" (Revista Iberoamericana de Inteligencia Artificial).

by the method of attributing certain mentalistic attitudes such as knowledge, belief —*information attitudes*, desire, intention, obligation, commitment —*pro-attitudes*, among others [5]. A well-known intentional system formal approach is the BDI architecture proposed by Rao and Georgeff [20, 21]. This model is based on the explicit representation of the agent's beliefs (B) —used to represent the state of the environment, its desires (D) —used to represent the motivations of the agent, and its intentions (I) —used to represent the goals of the agent. This architecture has evolved over time and it has been applied in several of the most significant multiagent applications developed up to now.

Modelling different intentional notions by means of several modalities (B, D, I) can be very complex if only one logical framework is used. In order to help in the design of such complex logical systems Giunchiglia et.al. [9] introduced the notion of *multi-context system* (MCS for short). This framework allows the definition of different formal components and their interrelation. In our case, we propose to use separate contexts to represent each modality and formalise each context with the most appropriate logic apparatus. The interactions between the components are specified by using inter-unit rules, called *bridge rules*. These rules are part of the deduction machinery of the system. This approach has been used by Sabater et.al. [22] and Parsons et.al. [19] to specify several agent architectures and particularly to model some classes of BDI agents [17]. Indeed one advantage of the MCS logical approach to agency modelling is that it allows for rather affordable computational implementation. For instance, a portion of the framework described in [17] is being now implemented using a prolog multi-threaded architecture [8].

The agent architectures proposed so far mostly deal with two-valued information. Although the BDI model developed by Rao and Georgeff explicitly acknowledges that an agent's model of the world is incomplete, by modelling beliefs as a set of worlds that the agent knows that it might be in, it makes no use of quantified information about how possible a particular world is to be the actual one. Neither does it allow desires and intentions to be quantified. We think that taking into consideration this graded information could improve the agent's performance. There are a few works that partially address this issue and emphasize the importance of graded models. Notably, Parsons and Giorgini [17] consider the belief quantification by using Evidence Theory. In their proposal, an agent is allowed to express its opinion on the reliability of the agents it interacts with, and to revise its beliefs when they become inconsistent. They set out the importance of quantifying degrees in desires and intentions, but this is not covered by their work. Lang et al. [14] present an approach to a logic of desires, where the notion of hidden uncertainty of desires is introduced. Desires are formalized to support a realistic interaction between the concepts of preference and plausibility (or normality), both represented by a pre-order relation over the sets of possible worlds. Other works deal with reasoning about intentions in uncertain domains, as the proposal of Schut et al. [24]. They present an efficient intention reconsideration for agents that interact in an uncertainty environment in terms of dynamics, observability, and non-determinism.

All the above mentioned proposals model partial aspects of the uncertainty related to mental notions involved in an agent's architecture. We present in this paper a general model for a graded BDI agent, specifying an architecture able to deal with the environment uncertainty and with graded mental attitudes. In this sense, belief degrees represent to what extent the agent believes a formula is true. Degrees of positive or negative desire allow the agent to set different levels of preference or rejection respectively. Intention degrees give also a preference measure but, in this case, modelling the cost/benefit trade off of reaching an agent's goal. Then, Agents having different kinds of behaviour can be modeled on the basis of the representation and interaction of these three attitudes.

This paper is organised as follows: in Section 2, we introduce multi-context systems and the general multivalued logic framework for the graded contexts. Sections 3, 4, and 5 present the mental units of the graded BDI model, that is the contexts for beliefs (BC), desires (DC), and intentions (IC). Section 6 outlines two functional contexts for planning (PC) and communication (CC). In Section 7, we deal with bridge rules, we illustrate the overall reasoning process in Section 8, and finally, we present some conclusions and future lines of work.

2 Graded BDI agent model

The architecture presented in this paper is inspired by the work of Parsons et.al. [17] about multi-context BDI agents. Multi-context systems were introduced by Giunchiglia et.al. [9] to allow different formal (logic) components to be defined and interrelated. The MCS specification of an agent contains three basic components: units or contexts, logics, and bridge rules, which channel the propagation of consequences among theories. Thus, an agent is defined as a group of interconnected units: $\langle \{C_i\}_{i \in I}, \Delta_{br} \rangle$, where each context $C_i \in \{C_i\}_{i \in I}$ is the tuple $C_i = \langle L_i, A_i, \Delta_i \rangle$ where L_i, A_i and Δ_i are the language, axioms, and inference rules respectively. They define the logic for the context and its basic behaviour as constrained by the axioms. When a theory $T_i \in L_i$ is associated with each unit, the implementation of a particular agent is complete. Δ_{br} can be understood as rules of inference with premises and conclusions in different contexts, for instance:

$$\frac{C_1 : \psi, C_2 : \varphi}{C_3 : \theta}$$

means that if formula ψ is deduced in context C_1 and formula φ is deduced in context C_2 then formula θ is added to context C_3.

The deduction mechanism of these systems is based on two kinds of inference rules, internal rules Δ_i inside each unit, and bridge rules Δ_{br} outside. Internal rules allow to draw consequences within a theory, while bridge rules allow to embed results from a theory into another [7].

We have *mental* contexts to represent beliefs (BC), desires (DC) and intentions (IC). We also consider two *functional* contexts: for Planning (PC) and Communication (CC). The Planner is in charge of finding plans to change the current world into another world, where some goal is satisfied, and of computing

the cost associated to the plans. The communication context is the agent's door to the external world, receiving and sending messages. In summary, the BDI agent model is defined as:

$$A_g = (\{BC, DC, IC, PC, CC\}, \Delta_{br})$$

Each context has an associated logic, that is, a logical language with its own semantics and deductive system. In order to represent and reason about graded notions of beliefs, desires and intentions, we decide to use a modal many-valued approach. In particular, we shall follow the approach developed by Hájek et al. in e.g. [12] and [10] where uncertainty reasoning is dealt with by defining suitable modal theories over suitable many-valued logics. The basic idea is the following. For instance, let us consider a Belief context where belief degrees are to be modeled as probabilities. Then, for each classical (two-valued) formula φ, we consider a modal formula $B\varphi$ which is interpreted as "φ is probable". This modal formula $B\varphi$ is then a *fuzzy* formula which may be more or less true, depending on the probability of φ. In particular, we can take as truth-value of $B\varphi$ precisely the probability of φ. Moreover, using a many-valued logic, we can express the governing axioms of probability theory as logical axioms involving modal formulae of the kind $B\varphi$. Then, the many-valued logic machinery can be used to reason about the modal formulae $B\varphi$, which faithfully respect the uncertainty model chosen to represent the degrees of belief.

In this proposal, for the mental contexts we choose the infinite-valued Lukasiewicz logic but another selection of many-valued logics may be done for each unit, according to the measure modeled in each case [1]. Therefore, in this kind of logical frameworks we shall have, besides the axioms of Lukasiewicz many-valued logic, a set of axioms corresponding to the basic postulates of a particular uncertainty theory. Hence, in this approach, reasoning about probabilities (or any other uncertainty models) can be done in a very elegant way within a uniform and flexible logical framework. The same many-valued logical framework may be used to represent and reason about degrees of desires and intentions, as will be seen in detail later on.

3 Belief Context

The purpose of this context is to model the agent's beliefs about the environment. In order to represent beliefs, we use modal many-valued formulae, following the above mentioned logical framework. We consider in this paper the particular case of using probability theory as the uncertainty model. Other models might be used as well by just modifying the corresponding axioms.

[1] The reason of using this many-valued logic is that its main connectives are based on the arithmetic addition in the unit interval [0, 1], which is what is needed to deal with additive measures like probabilities. Besides, Lukasiewicz logic has also the *min* conjunction and *max* disjunction as definable connectives, so it also allows to define a logic to reason about degrees of necessity and possibility.

3.1 The BC language

To reason about the credibility of crisp propositions, we define a language for belief representation, following Godo et al.'s [10], based on Lukasiewicz logic. In order to define the basic crisp language, we start from a classical propositional language L, defined upon a countable set of propositional variables PV and connectives $(\neg, \rightarrow)$, and extend it to represent actions. We take advantage of Dynamic logic which has been used to model agent's actions in [23] and [16]. These actions, the environment transformations they cause, and their associated cost must be part of any situated agent's beliefs set.

The propositional language L is thus extended to L_D, by adding to it action modalities of the form $[\alpha]$ where α is an action. More concretely, given a set Π_0 of symbols representing elementary actions, the set Π of plans (composite actions) and formulae L_D is defined as follows:

- $\Pi_0 \subset \Pi$ (elementary actions are plans)
- if $\alpha, \beta \in \Pi$ then $\alpha;\beta \in \Pi$, (the concatenation of actions is also a plan)
- if $\alpha, \beta \in \Pi$ then $\alpha \cup \beta \in \Pi$ (non-deterministic disjunction)
- if $\alpha \in \Pi$ then $\alpha^* \in \Pi$ (iteration)
- If A is a formula, then $A? \in \Pi$ (test)

- if $p \in PV$, then $p \in L_D$
- if $\varphi \in L_D$ then $\neg\varphi \in L_D$
- if $\varphi, \psi \in L_D$ then $\varphi \rightarrow \psi \in L_D$
- if $\alpha \in \Pi$ and $\varphi \in L_D$ then $[\alpha]\varphi \in L_D$.

The interpretation of $[\alpha]A$ is *"after the execution of α, A is true"*

We define a modal language BC over the language L_D to reason about the belief on crisp propositions. To do so, we extend the crisp language L_D with a fuzzy unary modal operator B. If φ is a proposition in L_D, the intended meaning of $B\varphi$ is that "φ is believable". Formulae of BC are of two types:

- *Crisp (non B-modal):* they are the (crisp) formulae of L_D, built in the usual way, thus, if $\varphi \in L_D$ then $\varphi \in BC$.
- *B-Modal:* they are built from elementary modal formulae $B\varphi$, where φ is crisp, and truth constants $\overline{r}$, for each rational $r \in [0,1]$, using the connectives of Lukasiewicz many-valued logic:
  - If $\varphi \in L_D$ then $B\varphi \in BC$
  - If $r \in Q \cap [0,1]$ then $\overline{r} \in BC$
  - If $\Phi, \Psi \in BC$ then $\Phi \rightarrow_L \Psi \in BC$ and $\Phi \& \Psi \in BC$ (where & and $\rightarrow_L$ correspond to the conjunction and implication of Lukasiewicz logic)

Other Lukasiewicz logic connectives for the modal formulae can be defined from $\&$, $\rightarrow_L$ and $\overline{0}$: $\neg_L\Phi$ is defined as $\Phi \rightarrow_L \overline{0}$, $\Phi \wedge \Psi$ as $\Phi\&(\Phi \rightarrow_L \Psi)$, $\Phi \vee \Psi$ as $\neg_L(\neg_L\Phi \wedge \neg_L\Psi)$, and $\Phi \equiv \Psi$ as $(\Phi \rightarrow_L \Psi)\&(\Psi \rightarrow_L \Phi)$.

Since in Lukasiewicz logic a formula $\Phi \rightarrow_L \Psi$ is 1-true iff the truth value of Ψ is greater or equal to that of Φ, modal formulae of the type $\overline{r} \rightarrow_L B\varphi$ express that the probability of φ is at least r. Formulae of the type $\overline{r} \rightarrow_L \Psi$ will be denoted as (Ψ, r).

3.2 Belief Semantics

The semantics for the language BC is defined, as usual in modal logics, using a Kripke structure. We have added to such structure a ρ function in order to represent the world transitions caused by actions, and a probability measure μ over worlds. Thus, we define a BC probabilistic Kripke structure as a 4-tuple $K = \langle W, e, \mu, \rho \rangle$ where:

- W is a non-empty set of possible worlds.
- $e : V \times W \rightarrow \{0, 1\}$ provides for each world a Boolean (two-valued) evaluation of the propositional variables, that is, $e(p, w) \in \{0, 1\}$ for each propositional variable $p \in V$ and each world $w \in W$. The evaluation is extended to arbitrary formulae in L_D as described below.
- $\mu : 2^W \rightarrow [0, 1]$ is a finitely additive probability measure on a Boolean algebra of subsets of W such that for each crisp φ, the set $\{w \mid e(\varphi, w) = 1\}$ is measurable [12].
- $\rho : \Pi_0 \rightarrow 2^{W \times W}$ assigns to each elementary action a set of pairs of worlds denoting world transitions.

Extension of e to L_D formulae:
e is extended to L using classical connectives and to formulae with action modalities –as $[\alpha] A$, by defining $\rho(\alpha; \beta) = \rho(\alpha) \circ \rho(\beta)$, $\rho(\alpha \cup \beta) = \rho(\alpha) \cup \rho(\beta)$, $\rho(\alpha^*) = (\rho(\alpha))^*$ (ancestral relation) and $\rho(\varphi?) = \{(w, w) \mid e(\varphi, w) = 1\}$, and setting $e([\alpha] A, w) = min \{e(A, w_i) \mid (w, w_i) \in \rho(\alpha)\}$. Notice that $e([\alpha] A, w) = 1$ iff the evaluation of A is 1 in all the worlds w' that may be reached through the action α from w.
Extension of e to B-modal formulae:
e is extended to B-modal formulae by means of Lukasiewicz logic truth-functions and the probabilistic interpretation of belief as follows:

- $e(B\varphi, w) = \mu(\{w' \in W \mid e(\varphi, w') = 1\})$, for each crisp φ
- $e(\overline{r}, w) = r$, for all $r \in Q \cap [0, 1]$
- $e(\Phi \& \Psi, w) = \max(e(\Phi) + e(\Psi) - 1, 0)$
- $e(\Phi \rightarrow_L \Psi, w) = \min(1 - e(\Phi) + e(\Psi), 1)$

Finally, the truth degree of a formula Φ in a Kripke structure $K = \langle W, e, \mu, \rho \rangle$ is defined as $\|\Phi\|^K = \inf_{w \in W} e(\Phi, w)$.

3.3 BC axioms and rules

As mentioned in Section 2, to set up an adequate axiomatization for our belief context logic we need to combine axioms for the crisp formulae, axioms of Lukasiewicz logic for modal formulae, and additional axioms for B-modal formulae according to the probabilistic semantics of the B operator. Hence, axioms and rules for the Belief context logic BC are as follows:

1. Axioms of propositional Dynamic logic for L_D formulae (see e.g. [11]).

2. Axioms of Lukasiewicz logic for modal formulae: for instance, axioms of Hájek's Basic Logic (BL) [12] plus the axiom: $\neg\neg\Phi \rightarrow \Phi$
3. Probabilistic axioms
   $B(\varphi \rightarrow \psi) \rightarrow_L (B\varphi \rightarrow B\psi)$
   $B\varphi \equiv \neg_L B(\varphi \wedge \neg\psi) \rightarrow_L B(\varphi \wedge \psi)$
   $\neg_L B\varphi \equiv B\neg\varphi$
4. Deduction rules for BC are: modus ponens, necessitation for $[\alpha]$ for each $\alpha \in \Pi$ (from φ derive $[\alpha]\varphi$), and necessitation for B (from φ derive $B\varphi$).

Deduction is defined as usual from the above axioms and rules and will be denoted by $\vdash_{BC}$. Notice that, taking into account Lukasiewicz semantics, the second *probabilistic axiom* corresponds to the finite additivity while the third one expresses that the probability of $\neg\varphi$ is 1 minus the probability of φ. Actually, one can show that the above axiomatics is sound and complete with respect to the intended semantics described in the previous subsection (cf. [12]). Namely, if T is a finite theory over BC and Φ is a (modal) formula, then $T \vdash \Phi$ iff $\|\Phi\|^K = 1$ in each BC probabilistic Kripke structure K model of T (i.e. K such that $\|\Psi\|^K = 1$ for all $\Psi \in T$).

4 Desire Context

In this context, we represent the agent's desires. Desires represent the *ideal* agent's preferences regardless of the agent's current perception of the environment and regardless of the cost involved in actually achieving them. We deem important to distinguish what is positively desired from what is not rejected. According to the works on bipolarity representation of preferences by Benferhat et.al. [2], positive and negative information may be modeled in the framework of possibilistic logic. Inspired by this work, we suggest to formalise agent's desires also as positive and negative. Positive desires represent what the agent would like to be the case. Negative desires correspond to what the agent rejects or does not want to occur. Both, positive and negative desires can be graded.

4.1 DC Language

The language DC is defined as an extension of a propositional language L by introducing two (fuzzy) modal operators D^+ and D^-. $D^+\varphi$ reads as "φ is positively desired" and its truth degree represents the agent's level of satisfaction would φ become true. $D^-\varphi$ reads as "φ is negatively desired" and its truth degree represents the agent's measure of disgust on φ becoming true. As in BC logic, we will use a modal many-valued logic to formalise graded desires. We use again Lukasiewicz logic as the base logic, but this time extended with a new connective Δ (known as Baaz's connective), considered also in [12]. For any modal Φ, if Φ has value < 1 then $\Delta\Phi$ gets value 0; otherwise, if Φ has value 1 then $\Delta\Phi$ gets value 1 as well. Hence $\Delta\Phi$ becomes a two-valued (Boolean) formula. Therefore, DC formulae are of two types:

- *Crisp (non modal):* formulae of L
- *Many-valued (modal):* they are built from elementary modal formulae $D^+\varphi$ and $D^-\varphi$, where φ is from L, and truth constants $\bar{r}$ for each rational $r \in [0,1]$:
  - If $\varphi \in L$ then $D^-\varphi, D^+\varphi \in DC$
  - If $r \in Q \cap [0,1]$ then $\bar{r} \in DC$
  - If $\Phi, \Psi \in DC$ then $\Phi \rightarrow_L \Psi \in DC$ and $\Phi \& \Psi \in DC$

As in BC, $(D\psi, \bar{r})$ denotes $\bar{r} \rightarrow_L D\psi$.

In this context the agent's preferences will be expressed by a theory T containing quantitative expressions about positive and negative preferences, like $(D^+\varphi, \alpha)$ or $(D^-\psi, \beta)$, as well as qualitative expressions like $D^+\psi \rightarrow_L D^+\varphi$ (resp. $D^-\psi \rightarrow_L D^-\varphi$), expressing that φ is at least as preferred (resp. rejected) as ψ. In particular $(D^+\phi_i, 1) \in T$ means that the agent has maximum preference in ϕ_i and is fully satisfied if it is true. While $(D^+\phi_j, \alpha) \notin T$ for any $\alpha > 0$ means that the agent is indifferent to ϕ_j and the agent doesn't benefit from the truth of ϕ_j. Analogously, $(D^-\psi_i, 1) \in T$ means that the agent absolutely rejects ϕ_i and thus the states where ψ_i is true are totally unacceptable. $(D^-\psi_j, \beta) \notin T$ for any $\beta > 0$ simply means that ψ_j is not rejected, the same applies to the formulae not explicitly included in T.

4.2 Semantics for DC

The degree of positive desire for (or level of satisfaction with) a disjunction of goals $\varphi \vee \psi$ is taken to be the minimum of the degrees for φ and ψ. Intuitively if an agent desires $\varphi \vee \psi$ then it is ready to accept the situation where the less desired goal becomes true, and hence to accept the minimum satisfaction level produced by one of the two goals. In contrast the satisfaction degree of reaching both φ and ϕ can be strictly greater than reaching one of them separately. These are basically the properties of the *guaranteed possibility* measures (see e.g. [1]). Analogously, we assume the same model for the degrees of negative desire or rejection, that is, the rejection degree of $\varphi \vee \phi$ is taken to be the minimum of the degrees of rejection for φ and for ψ separately, while nothing prevents the rejection level of $\varphi \wedge \psi$ be greater than both.

The DC models are Kripke structures $M_D = \langle W, e, \pi^+, \pi^- \rangle$ where W and e are defined as in the BL semantics and π^+ and π^- are preference distributions over worlds, which are used to give semantics to positive and negative desires:

- $\pi^+ : W \rightarrow [0,1]$ is a distribution of positive preferences over the possible worlds. In this context $\pi^+(w) < \pi^+(w')$ means that w' is more preferred than w.
- $\pi^- : W \rightarrow [0,1]$ is a distribution of negative preferences over the possible worlds: $\pi^-(w) < \pi^-(w')$ means that w' is more rejected than w.

We impose a consistency condition: $\pi^-(w) > 0$ implies $\pi^+(w) = 0$, that is, if w is rejected to some extent, it cannot be desired. And conversely. The truth evaluation e is extended to the non-modal formulae in the usual (classical) way.

The extension to modal formulae uses the preference distributions for formulae $D^-\varphi$ and $D^+\varphi$, and for the rest of modal formulae by means of Lukasiewicz connectives, as in BC semantics, plus the unary connective Δ. The evaluation of modal formulae only depends on the formula itself –represented in the preference measure over the worlds where the formula is true– and not on the actual world where the agent is situated:

- $e(D^+\varphi, w) = \inf\{\pi^+(w') \mid e(\varphi, w') = 1\}$
- $e(D^-\varphi, w) = \inf\{\pi^-(w') \mid e(\varphi, w') = 1\}$
- $e(\Delta\Phi, w) \begin{cases} 1, \text{ if } e(\Phi, w) = 1 \\ 0, \text{ otherwise.} \end{cases}$

As usual, by convention we take $\inf \emptyset = 1$ and thas $e(D^+\bot, w) = e(D^-\bot, w) = 1$ for all $w \in W$.

4.3 DC Axioms

In a similar way as in BC, to axiomatize the logical system DC we need to combine classical logic axioms for non-modal formulae with Lukasiewicz logic axioms extended with Δ for modal formulae. Also, additional axioms characterizing the behaviour of the modal operators D^+ and D^- are needed. Hence, we define the axioms and rules for the DC logic as follows:

1. Axioms of classical logic for the non-modal formulae.
2. Axioms of Lukasiewicz logic with Δ (cf. [12]) for the modal formulae.
3. Axioms for D^+ and D^- over Lukasiewciz logic:
   $D^+(A \vee B) \equiv D^+A \wedge D^+B$
   $D^-(A \vee B) \equiv D^-A \wedge D^-B$
   $\neg_L\Delta(D^+A \wedge D^-A) \rightarrow \neg_L(\nabla D^-A \& \nabla D^+A)$, where ∇ is $\neg_L\Delta\neg_L$[2].
   $D^+(\bot)$
   $D^-(\bot)$
4. Rules are: modus ponens, necessitation for Δ, and introduction of D^+ and D^- for implications: from $A \rightarrow B$ derive $D^+B \rightarrow_L D^+A$ and $D^-B \rightarrow_L D^-A$.

Notice that the two first axioms in item (3) define the behaviour of D^- and D^+ with respect to disjunctions, while the third axiom establishes that it is not possible to have at the same time positive and negative desires over the same formula except if the formula is a contradiction. In that case notice that the antecedent of the axiom becomes false. Finally, the two inference rules state that the degree of desire is monotonically decreasing with respect to logical implication. This axiomatics is correct with respect to the above defined semantics, and the conjecture is that it is complete too.

[2] Notice that $e(\nabla\Phi, w) = 1$ if $e(\Phi, w) > 0$, and $e(\nabla\Phi, w) = 0$ otherwise.

5 Intention Context

In this context, we represent the agent's intentions. We follow the model introduced by Rao and Georgeff [20, 21], in which an intention is considered a fundamental pro-attitude with an explicit representation. Intentions, as well as desires, represent the agent's preferences. However, we consider that intentions cannot depend just on the benefit, or satisfaction, of reaching a goal φ –represented in $D^{+}\varphi$, but also on the world's state w and the cost of transforming it into a world w_i where the formula φ is true. By allowing degrees in intentions we represent a measure of the cost/benefit relation involved in the agent's actions towards the goal. The positive and negative desires are used as pro-active and restrictive tools respectively, in order to set intentions. Note that intentions depend on the agent's knowledge about the world, which may allow –or not– the agent to set a plan to change the world into a desired one. Thus, if in a theory T we have the formula $I\psi \rightarrow_L I\varphi$ then the agent may try φ before ψ and it may not try ϕ if $(I\phi, \delta)$ is a formula in T and $\delta < Threshold$. This situation may mean that the benefit of getting ϕ is low or the cost is high.

5.1 IC Language

We define its syntax in the same way as we did with BC (except for the dynamic logic part), starting with a basic language L and incorporating a modal operator I. We use Lukasiewicz multivalued logic to represent the degree of the intentions. As in the other contexts, if the degree of $I\varphi$ is δ, it may be considered that the truth degree of the expression "φ is intended" is δ. The intention to make φ true must be the consequence of finding a feasible plan α, that permits to achieve a state of the world where φ holds.

The value of $I\varphi$ will be computed by a bridge rule (see (3) in next Section 7), that takes into account the benefit of reaching φ and the cost, estimated by the Planner, of the possible plans towards it.

5.2 Semantics and axiomatization for IC

The semantics defined in this context shows that the value of the intentions depends on the formula intended to bring about and on the benefit the agent gets with it. It also depends on the agent's knowledge on possible plans that may change the world into one where the goal is true, and their associated cost. This last factor will make the semantics and axiomatization for IC somewhat different from the presented for positive desires in DC.

The models for IC are Kripke structures $K = \langle W, e, \{\pi_w\}_{w \in W} \rangle$ where W and e are defined in the usual way, and for each $w \in W$, $\pi_w : W \rightarrow [0, 1]$ is a possibility distribution where $\pi_w(w') \in [0, 1]$ is the degree on which the agent may try to reach the state w' from the state w.

The truth evaluation $e : V \times W \rightarrow \{0, 1\}$ is extended to the non-modal formulae in the usual way. It is extended to modal formulae using Lukasiewicz semantics as $e(I\varphi, w) = N_w(\{w' \mid e(\varphi, w') = 1\})$, where N_w denotes the necessity

measure associated to the possibility distribution π_w, defined as $N_w(S) = \inf\{1 - \pi_w(s) \mid s \notin S\}$. A sound and complete axiomatics for the I operator, is defined in a similar way as for the previous mental operators but now taking the axioms corresponding to necessity measures (cf. [12]), that is, the following axioms:

1. Axioms of classical logic for the non-modal formulae.
2. Axioms of Lukasiewicz logic for the modal formulae.
3. Axioms for I over Lukasiewciz logic:
   $I(\varphi \to \psi) \to (I\varphi \to I\psi)$
   $\neg I(\bot)$
   $I(\varphi \wedge \psi) \equiv (I\varphi \wedge I\psi)$
4. Deduction rules are modus ponens and necessitation for I (from φ derive $I\varphi$).

6 Planner and Communication Contexts

The nature of these contexts is functional. The Planner Context (PC) has to build plans which allow the agent to move from its current world to another, where a given formula is satisfied. This change will indeed have an associated cost according to the actions involved. Within this context, we propose to use a first order language restricted to Horn clauses (PL), where a theory of planning includes the following special predicates:

- *action*(α, P, A, c_α) where $\alpha \in \Pi_0$ is an elementary action, $P \subset PL$ is the set of preconditions; $A \subset PL$ are the postconditions and $c_\alpha \in [0, 1]$ is the normalised cost of the action.
- *plan*(φ, α, P, A, c_α, r) where $\alpha \in \Pi$ is a composite action representing the plan to achieve φ, P are the pre-conditions of α, A are the post-conditions $\varphi \in A$, c_α is the normalized cost of α and r is the belief degree (> 0) of actually achieving φ by performing plan α. We assume that only one instance of this predicate is generated per formula.
- *bestplan*(φ, α, P, A, c_α, r) similar to the previous one, but only one instance with the best plan is generated.

Each plan must be feasible, that is, the current state of the world must satisfy the preconditions, the plan must make true the positive desire the plan is built for, and cannot have any negative desire as post-condition. These feasible plans are deduced by a bridge rule among the BC, DC and PC contexts (see (2) in the next Section 7).

The communication unit (CC) makes it possible to encapsulate the agent's internal structure by having a unique and well-defined interface with the environment. This unit also has a first order language restricted to Horn clauses. The theory inside this context will take care of the sending and receiving of messages to and from other agents in the Multi Agent society where our graded BDI agents live. Both contexts use resolution as a deduction method.

7 Bridge Rules

For our BDI agent model, we define a collection of basic bridge rules to set the interrelations between contexts. These rules are illustrated in figure 1. In this section we comment the most relevant ones.

The agent's knowledge about the world's state and about actions that change the world, is introduced from the belief context into the Planner as first order formulae $\lceil . \rceil$:

$$\frac{B : B\varphi}{P : \lceil B\varphi \rceil} \tag{1}$$

Then, from the positive desires, the beliefs of the agent, and the possible transformations using actions, the Planner can build plans. Plans are generated from actions, to fulfill positive desires, but avoiding negative desires. The following bridge rule among D, B, and P contexts does this:

$$\frac{\begin{array}{c} D : \nabla(D^{+}\varphi), D : (D^{-}\psi, \mathit{threshold}), P : \mathit{action}(\alpha, P, A, c), \\ B : (B([\alpha]\varphi), r), B : B(A \rightarrow \neg\psi) \end{array}}{P : \mathit{plan}(\varphi, \alpha, P, A, c, r)} \tag{2}$$

As we have previously mentioned, the intention degree trades off the benefit and the cost of reaching a goal. There is a bridge rule that infers the degree of $I\varphi$ for each plan α that allows to achieve the goal. This value is deduced from the degree of $D^{+}\varphi$ and the cost of a plan that satisfies desire φ. This degree is calculated by function f as follows:

$$\frac{D : (D^{+}\varphi, d), P : \mathit{plan}(\varphi, \alpha, P, A, c, r)}{I : (I\varphi, f(d, c, r))} \tag{3}$$

Different functions model different individual behaviours. For example, if we consider an *equilibrated agent*, the degree of the intention to bring about φ, under full belief in achieving φ after performing α, may depend equally on the satisfaction that it brings the agent and in the cost—considering the complement to 1 of the normalised cost. So the function might be defined as

$$f(d, c, r) = r(d + (1 - c))/2.$$

In fact, given the plan P for the goal φ, with desire level d and(normalized) cost c, we can think of $u = (d + (1 - c))/2$ as the utility of reaching φ by means of the plan P. The intention degree as computed above is thennothing but $r \cdot u$, that is, the utility u multiplied by the probability r of reaching φ after the plan is executed. This is actually the *expected utility* of reaching φ by means of the plan P if one considers a utility value of 0 when the plan P does not reach φ.

In BDI agents, bridge rules have been also used to determine the relationship between the mental attitudes and the actual behaviour of the agent. Well-established sets of relations for BDI agents have been identified [21]. If we use the *strong realism* model, the set of intentions is a subset of the set of desires,

which in turn is a subset of the beliefs. That is, if an agent does not believe something, it will neither desire it nor intend it [20]:

$$\frac{B : \neg B\psi}{D : \neg D\psi} \text{ and } \frac{D : \neg D\psi}{I : \neg I\psi} \tag{4}$$

We also need bridge rules to establish the agent's interactions with the environment, meaning that if the agent intends φ at degree i_{max}, where i_{max} is the maximum degree of all the intentions, then the agent will focus on the plan -bestplan- that allows the agent to reach the most intended goal:

$$\frac{I : (I\varphi, i_{max}), P : bestplan(\varphi, \alpha, P, A, c_\alpha, r)}{C : C(does(\alpha))} \tag{5}$$

Through the comunication unit the agent perceives all the changes in the enviroment that are introduced by the following bridge rule in the belief context:

$$\frac{C : \beta}{B : B\beta} \tag{6}$$

Figure 1 shows the graded BDI agent proposed with the different contexts and the bridge rules relating them.

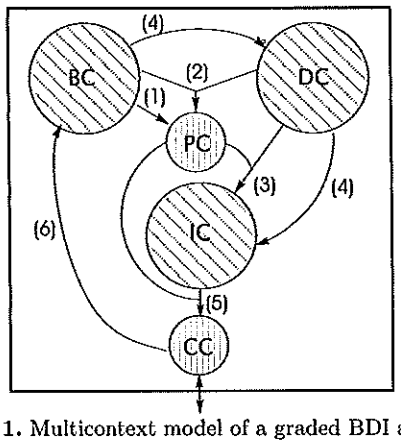

Fig. 1. Multicontext model of a graded BDI agent

8 Example of a graded BDI Agent for tourism

Suppose we want to instruct our travel agent to look for a one-week holiday destination package. We instruct the agent with two desires, first and more important, we want to rest, and second we want to visit new places (visitNP). We restrict its exploration range as we do not want to travel more than 1000 kms from Rosario, where we live. To propose a destination (plan) the agent will have to take into account the benefit (with respect to rest and to visitNP) and

the cost of the travel. The agent will consult with a travel agency that will give a number of plans, that conveniently placed in the planner context will determine the final set of proposals. In this scenario we have the following theories in the BC, DC, and PC contexts (IC has no initial theory):

D context: The agent has the following positive and negative desires:

- $(D^{+}(rest), 0.8)$
- $(D^{+}(visitNP), 0.7)$
- $(D^{+}(rest \wedge visitNP), 0.9)$
- $(D^{-}(distance > 1000km), 0.9)$

B context: This theory contains knowledge about the relationship between possible actions the agent can take and formulae made true by their execution. In this case, actions would be *traveling* to different destinations. For this example we consider only six destinations:

$$\Pi_0 = \{CarlosPaz, Cumbrecita, Bariloche, VillaGesell, MardelPlata, PtoMadryn\}.$$

Then, we represent the agent's beliefs about visiting new places and resting. In particular, we may consider the degree of $B([\alpha]visitNP)$ as the probability of *visitNP* after traveling to α. According to the places we know in each destination and the remaining places to visit in each destination, we give our travel agent the following beliefs:

- (B([Cumbrecita]visitNP), 1)
- (B([Carlos Paz]visitNP), 0.3)
- (B([Bariloche]visitNP), 0.7)
- (B([Villa Gesell]visitNP), 0.6)
- (B([Mar del Plata]visitNP), 0.3)
- (B([Pto Madryn]visitNP), 1)

The agent needs to assess also beliefs about the possibility that a destination offers to rest. In this case the degree of $B([\alpha]Rest)$ is interpreted as the probability of resting in α. These beliefs are determined by the characteristics of the destination —beach, mountains, big or a small city, etc— and taking into account our personal views:

- (B([Cumbrecita]Rest), 1)
- (B([Carlos Paz]Rest), 0.8)
- (B([Bariloche]Rest), 0.6)
- (B([Villa Gesell]Rest), 0.8)
- (B([Mar del Plata]Rest), 0.5)
- (B([Pto Madryn]Rest), 0.7)

We assume here that, for each action α, the positive desires are stochastically independent, so we add to BC an appropriate inference rule:

$$\frac{(B[\alpha]Rest, r), (B[\alpha]visitNP, s)}{(B[\alpha](Rest \wedge visitNP), r \cdot s)}$$

P Context A series of elementary actions:

- action (Cumbrecita, {cost = 800},{dist =500 km}, 0.67)
- action (Carlos Paz, {cost = 500},{dist = 450 km}, 0.42)
- action (Bariloche, {cost = 1200},{dist = 1800 km},1)
- action (Pto Madryn, {cost = 1000},{dist =1700 km}, 0.83)
- action (Villa Gessell, {cost = 700},{dist =700 km}, 0.58)
- action (Mar del Plata, {cost = 600},{dist =850 km}, 0.5)

Once these theories are defined the agent is ready to reason in order to determine which Intention to adopt and which plan is associated with that intention. We follow give a brief schema of the different steps in this process:

1. *The desires are passed from DC to PC.*
2. *Within PC plans for each desire are found.*
   Starting from the positive desires the planner looks for a set of different destination plans, taking into consideration the beliefs of the agent about the possibilities of satisfying the goals rest and visitNP through the different actions. Using the restriction introduced by the negative desire: $(D^{-}(dist > 1000km), 0.9)$ the planner rejects plans to Bariloche and to Pto Madryn, because their post-conditions make true $(dist > 1000km)$ which is strongly rejected (0.9). Therefore, using the bridge rule (2), plans are generated for each desire. For instance, for the most preferred desire, i.e. $rest \wedge visitNP$ the following plans are generated:

   $plan(rest \wedge visitNP, Cumbrecita, \{cost = 800\}, \{dist = 500km\}, 0.67, 1)$
   $plan(rest \wedge visitNP, CarlosPaz, \{cost = 500\}, \{dist = 450km\}, 0.42, 0.24)$
   $plan(rest \wedge visitNP, VillaGessell, \{cost = 700\}, \{dist = 700km\}, 0.58, 0.48)$
   $plan(rest \wedge visitNP, MardelPlata, \{cost = 600\}, \{dist = 850km\}, 0.5, 0.15)$

3. *The plans determine the degree of intentions.*
   Using bridge rule (3) and the function f proposed for an *equilibrated* agent the I context calculates the intention degree for the different destinations. Since f is monotonically increasing with respect to d, it is enough to consider the most preferred desired, i.e. $rest \wedge visitNP$. Hence, $rest \wedge visitNP$ is preferred to a degree 0.9, using $f(d, b, c) = b(0.9 + (1 - c))/2$ we successively have for $\alpha \in \{Cumbrecita, CarlosPaz, VillaGessell, MardelPlata\}$:

   $(I(rest \wedge visitNP), 0.615)$,
   $(I(rest \wedge visitNP), 0.1776)$,

$(I(rest \wedge visitNP), 0.3168)$,
$(I(rest \wedge visitNP), 0.105)$.
We get a maximal degree of intention for $rest \wedge visitNP$ by the plan *cumbrecita*, of 0.615.

4. *A plan is adopted.*
Finally, by means of bridge rule (5), the action $\alpha = Cumbrecita$ is selected and passed to the Communication context CC.

9 Conclusions and Future Work

This paper has presented a BDI agent model that allows to explicitly represent the uncertainty of beliefs, desires and intentions. This graded architecture is specified using multicontext systems and is general enough to be able to specify different types of agents. In this work we have used a different context for each attitude: Belief, Desire and Intention. We used a specific logic for each unit, according to the attitude represented. The Lukasiewicz multivalued logic is the framework chosen to formalise the degrees and we added the corresponding axiomatic in order to represent the uncertainty behaviour as probability, necessity and possibility. Other measures of uncertainty might be used in the different units by simply changing the corresponding axiomatic. Adding concrete theories to each context, particular agents may be defined using our context blueprints. The agent's behaviour is then determined by the different uncertainty measures of each context, the specific theories established for each unit, and the bridge rules. An issue of current research is to look for possible alternative axiomatic modelings of desires and intentions, and their implications in the bridge rules which deal with them, and check how they can also influence the agent's behavior. Besides, the model introduced, based on a multicontext specification, can be easily extended to include other mental attitudes.

As for future work, we are considering two directions. On the one hand we want to extend our multicontext agent model to a multiagent scenario. We plan to do this by introducing a *social context* in the agent architecture to deal with all aspects of social relations with other agents. In particular to equip this social context with a good logical model of trust is very important to allow the agent to infer beliefs from other agents' information. Interesting models of trust are Liau's logic of Belief, Information and Trust (BIT) [15] in the extension of this model described in [4] in this volume.

On the other hand, from an computational point of view, our idea is to implement each unit as prolog thread, equipped with its own meta-interpreter. The meta-interpreter purpose will be to manage inter-thread (inter-context) communication, i.e. all processes regarding bridge rule firing and assertion of bridge rule conclusions into the corresponding contexts. This implementation will support both, the generic definition of graded BDI agent architectures and the specific instances for particular types of agents. The implementation will also allow us to experiment and validate the formal model presented.

Acknowledgments Lluis Godo acknowledges partial support by the Spanish project MULOG, TIN2004-07933-C03-01, and Carles Sierra acknowledges partial support by the Spanish project WEBI2, TIC2003-08763-C02-00.

References

1. Benferhat S., Dubois D., Kaci S. and Prade, H. Bipolar Possibilistic Representations. *Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence (UAI 2002)*: pages 45-52. Morgan Kaufmann 2002.
2. Benferhat S., Dubois D., Kaci S. and Prade, H. Bipolar representation and fusion of preferences in the possilistic Logic framework. *In Proceedings of the 8th International Conference on Principle of Knowledge Representation and Reasoning (KR-2002)*, pages 421-448, 2002.
3. Cimatti A. and Serafini L. Multi-Agent Reasoning with Belief Contexts: the Approach and a Case Study. In M. Wooldridge and N. R. Jennings, editors, *Intelligent Agents: Proceedings of 1994 Workshop on Agent Theories, Architectures, and Languages*, num-ber 890 in Lecture Notes in Computer Science, pages 71-5. Springer Verlag, 1995.
4. Dastani M., Herzig A., Hulstijn J. and van der Torre L. Inferring Trust. Proceedings of the *5th International Workshop on Computational Logic in Multiagent Systems* (CLIMA V), Leite J. and Torroni P. editors. This volume.
5. Dennet, D. C. *The Intentional Stance.* MIT Press, Cambridge, MA, 1987.
6. Esteva, F., Garcia, P. and Godo L. Relating and extending semantical approaches to possibilistic reasoning. *International Journal of Approximate Reasoning*, 10:311-344, 1994.
7. Ghidini C. and Giunchiglia F. Local Model Semantics, or Contextual Reasoning = Locality + Compatibility *Artificial Intelligence*,127(2):221-259, 2001.
8. Giovannucci A. Towards Multi-Context based Agents Implementation. IIIA-CSIC Research Report, in preparation.
9. Giunchiglia F. and Serafini L. Multilanguage Hierarchical Logics (or: How we can do without modal logics) *Journal of Artificial Intelligence*, vol.65, pp. 29-70, 1994.
10. Godo, L., Esteva, F. and Hajek, P. Reasoning about probabilities using fuzzy logic. *Neural Network World*, 10:811–824, 2000.
11. Goldblatt R. *Logics of Time and Computation*, CSLI Lecture Notes 7, 1992.
12. Hájek, P. *Metamathematics of Fuzzy Logic*, volume 4 of Trends in Logic. Kluwer, 1998.
13. Jennings N.R. On Agent-Based Software Engineering. *Artificial Intelligence* 117(2), 277-296, 2000.
14. Lang J., van der Torre, L. and Weydert E. Hidden Uncertainty in the Logical Representation of Desires *International Joint Conference on Artificial Intelligence, IJCAI 03*, Acapulco, Mexico, 2003.
15. Liau C. J., Belief, Information Acquisition, and Trust in Multiagent Systems - a modal formulation. *Artificial Intelligence* 149, 31-60, 2003.
16. Meyer J. J. Dynamic Logic for Reasoning about Actions and Agents. *Workshop on Logic-Based Artificial Intelligence*, Washington, DC, June 14–16, 1999
17. Parsons, S., Sierra, C. and Jennings N.R. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3): 261-292, 1998.
18. Parsons, S. And Giorgini P. On using degrees of belief in BDI agents. *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Paris, 1998.

19. Parsons,S., Jennings,N.J., Sabater,J. and Sierra C. Agent Specification Using Multi-context Systems. *Foundations and Applications of Multi-Agent Systems 2002:* 205-226, 2002.
20. Rao, A. And Georgeff M. Modeling Rational Agents within a BDI-Architecture. *In proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR-92)*, pages 473-484 (ed R. Fikes and E. Sandewall), Morgan Kaufmann, San Mateo, CA, 1991.
21. Rao, A. and Georgeff M. BDI agents: From theory to practice. *In proceedings of the 1st International Conference on Multi-Agents Systems*, pp 312-319, 1995.
22. Sabater,J., Sierra, C., Parsons,S. and Jennings N. R. Engineering executable agents using multi-context systems. *Journal of Logic and Computation*12(3): 413-442 (2002).
23. Sierra, C., Godo,L., López de Màntaras, R. and Manzano Descriptive Dynamic Logic and its Application to Reflective Architectures. *Future Generation Computer Systems*, 12, 157-171, 1996.
24. Schut, M., Wooldridge, M. and Parsons S. Reasoning About Intentions in Uncertain Domains Symbolic and Quantitative Approaches to Reasoning with Uncertainty. *6th ECSQARU 2001, Proceedings*, pages 84-95, Toulouse, France, 2001.
25. Wooldridge, M and Jennings N.R. Intelligent Agents: theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152, 1995.