RESEARCH WORK REPORT

Extension of the BDI Agent Model: Representing and Reasoning using Graded Attitudes

By Ana Casali

Directed by: Lluis Godo and Carles Sierra
Institut d'Investigació en Intel·ligència Artificial
IIIA - CSIC

PhD Program in Information Technologies
Department of Electronics, Computer Sc. and Automatic Control
UNIVERSITY OF GIRONA
Girona, Spain - 2005

Contents

Chapter 1

Introduction

The aim of this Report is to present a research work I have done under the direction of Lluis Godo and Carles Sierra. This work undertakes an extension of the BDI agent architecture in order to incorporate the representation of uncertainty in the beliefs, to allow the expression of graded positive and negative desires, and graded intentions too. How did we decide to do this? Particularly, I have studied some years ago different approaches to approximate reasoning and how these models help to make knowledge based systems more flexible and useful for real applications. Now, in the frame of multiagents systems, in a distributed and complex platform of autonomous, proactive, reactive and social agents; I asked myself how the ideas underlying approximate reasoning could be extended and applied to these systems.

Following this motivation, we found interesting a paper of Parsons and Giorgini [58] where they make a first approach to a graded BDI model, including only the representation of uncertainty in the beliefs, and leaving the general graded model as an open research work. This "open door" to future work encouraged us to follow this research's direction.

There are other contributions which treat different aspects of agents that reason under uncertain or imprecision in dynamics and complex environments. The extension of the BDI model proposed in this work is only one step in this field, but we think it is an important one, because of the relevance of the BDI architecture and because some of the ideas may be adapted to other agent's architectures.

We consider that making the BDI architecture more flexible, will allow us to design and develop agents capable to have a better performance in uncertain and dynamic environments, serving to other agents (human or not) that may have a set of graded motivations.

This report is structured in the following way. In Chapter 2 we present the State of the Art about several aspects related to the problem we decided to tackle. It is divided in Fundamental Concepts and Related Works. The first Section includes the main concepts that support this research work as: Agents and Multiagents systems, particularly stressing some of the challenges they face in the next future; a review on Agent's theories and Architectures, with special attention to the BDI model; and finally, the Multi-context systems are presented paying special attention to its approach to agent's specification and engineering. In the second Section, we present the State of the Art of Related Works, we include here the review of some logics important to our work, such as: the many-valued approach to uncertainty and some logics of preference. Then we analyze a few works about extensions to the BDI model where some of the attitudes —belief, desires or intentions— where considered to be graded or ordered. We conclude the State of the Art with some observations and remarks about this review, that help us to outline our work. In Chapter 3, the general framework of the "Graded BDI agent model" is presented and then, in the consecutive Sections its fundamentals components as the different Contexts and the Bridge Rules are formalized. Later on, in order to give a view of how this model works, we show an example of a travel assistant agent.

Finally, some conclusions of the developed model are taken and we draft some work for the next future. This planned work is described in the Thesis Project, in Chapter 4.

Chapter 2

State of the Art

2.1 Fundamental Concepts

2.1.1 Agents and multiagents systems

In the recent past, an increasing number of multiagent systems (MAS) have been designed and implemented to engineer complex distributed systems [74, 44].

Computer applications play an increasingly important part of our everyday life. They are becoming more tightly connected each other in large networks and with humans through user-interfaces. Much of these systems are too complex to be completely characterized and precisely described; hence, these applications are hard to solve in centralized computing technology. Also, several of these systems are inherently distributed in the sense that the data and information to be processed is distributed geographically, temporarily or are structured into cluster whose access and use requires different capabilities [72]. Since that, multiagent systems stand as a promising way to understand, manage and use these distributed, large-scale, dynamic, open and heterogeneous computing and information systems.

From another point of view multiagent systems offers a natural way to understand and characterize intelligent systems. Intelligence and interaction are deeply coupled and these systems allows to reflect this insight. Several researchers argue that intelligent behavior is not disembodied, but is a product of the interaction the agent maintains with its environment. Under this conception multiagent systems stand as a new approach to Artificial Intelligence [65].

With the spread of multiagent systems there are an increasing number of projects and researchers involved in related fields. They have taken important co-ordination actions for agent based computing as for example AgentLink III, an European network of researchers and developers with a common interest in agent technology [39]. There are also several websites to provide information resource on intelligent agents, examples of these are Agentcities [42], AgentTechnologies [41] and Agentland [40], among others.

The agent society holds that there are some application domains where agent technologies will play a crucial role in the near future, including: ambient intelligence, grid computing, electronic business, the semantic web, bioinformatics, monitoring and control, resource management, space, military and manufacturing applications. They point out that the impact of agent technologies in application domains such as these will occur firstly, as a metaphor for the design of complex, distributed computational systems; secondly, as a source of technologies for such computing systems, and thirdly, as models of complex real-world systems [52].

In order to achieve the full potential of agent approaches and technologies there are a number of broad technological challenges for the next future. In the Agent Technology Roadmap of the AgentLink network, it is recommended that research and development resources may be focused along several key directions. Among these, to strengthen links with other areas of Computer Science working on different problems, like the uncertainty community in AI. In particular, M. Luck et al. in [52] recommend:

- Build bridges, especially to the artificial life, robotics, Uncertainty in AI, logic programming and the traditional mathematical modeling communities.
- Develop agent-based systems using hybrid approaches.
- Develop metrics to assess the relative strengths and weakness of different approaches.

The development of the work presented in this Report can be placed within these mention directions.

2.1.2 Agent's Theories and Architectures

In order to give multiagent systems a formal support, several researchers have proposed diverse theories and architectures for agents. Agent theories are essentially specifications of agents' behavior expressed as the properties that agents should have. A formal representation of the properties helps the designer to reason about the expected behavior of the system.

Agent's architectures can be thought of as software engineering models of agents and represent a middle point between specification and implementation. They identify the main functions that ultimately determine the agent's behavior and define the interdependencies that exist among them. A relevant review of the work done about agent's theories and architectures is due to M. Wooldridge and N. Jennings in [73]

Agent's theories based on an *intentional stance* are among the most common ones. These are based on a folk psychology by which human behavior is predicted and explained through the attribution of attitudes. For example, when explaining human activity, it is often useful and common to make statements such as the following:

Jorge took his coat because he *believed* it was going to be cold.

Peter worked hard because he *wanted* to save money.

In these examples, Jorge's and Peter's behaviors can be explained in terms of their attitudes, such as believing and wanting.

The philosopher Dennet has coined the term Intentional system to describe entities "whose behavior can be predicted by the method of attributing certain mentalistic attitudes such as belief, desires and rational acumen" [20]. Dennet also identifies different grades of intentional systems: a first-order intentional system has believes and desires (etc.) but no beliefs and desires about beliefs and desires. A second-order intentional system is a more sophisticated; it has beliefs and desires (and possibly other intentional states) about beliefs and desires (and other intentional states), both those of others and its own.

When the underlying system process is well known and understood, there is no reason to take an intentional stance, but this is not the case in many of the applications to real systems. The intentional notions are abstraction tools, which provide with a convenient and familiar way of describing, explaining, and predicting the behavior of complex systems. Considering that an agent is a system that is conveniently described by the intentional stance, it is worth to weigh up which attitudes are appropriate for representing agents. The two most important categories are *information attitudes* –knowledge, belief– and *pro-attitudes* –desire, intention, obligation, commitment, choice, among others.

Information attitudes are related to the knowledge that the agent has about the world, whereas pro-attitudes are those that in some way guide the agents actions. The attitudes of both categories are closely related and much of the work in agent theory is concerned with clearing up the relationships between them. Although there is no total agreement on which combination of attitudes is the most appropriate to characterize an agent, it seems reasonable that an agent must be represented in terms of at least one information-attitude and one pro-attitude.

There are various formalisms that focused on just one aspect of agency (i.e., beliefs, goals, intentions, etc.) but it is expected that a realistic agent theory will be represented in a logical framework that combines these different components. It is expected that a realistic agency theory will be represented in a logical framework that define how the attributes of agency are related; how an agents cognitive state changes over time; how the environment affects the agents believes; and how the agents information and pro-attitudes lead it to perform actions [73].

Considering now the area of agent architectures, P. Maes defines an agent architecture as:

> ' A particular methodology for building agents. It specifies how...the agent can be decomposed into the construction of a set of component modules and how these modules should be interact. The total set of modules and their interactions has to provide an answer to the question of how the sensor data and the current mental state of the agent determine the actions...and future mental state of the agent. An architecture encompasses techniques and algorithms that support this methodology.' (Maes [51], p115)

Making specific commitments about the internal structure and operation of agents, we have a distinct class of agents. There exists different proposal for the classification of agent's architectures. Following the classification defined by Wooldridge in [72, 76], we consider four classes of concrete architectures for intelligent agents :

1. *Logic based architectures(deductive agents)*
2. *Reactive architectures (reactive agents)*
3. *Layered architectures (hybrid agents)*
4. *Practical reasoning architectures (BDI agents)*

In the rest of this Subsection we outline the main characteristics of each kind of architecture and in the following Section we present the BDI model in more detail.

1. Logic-Based Architectures

A classical approach to building agents is following the traditional approach to building artificial intelligent systems. This paradigm suggest that the intelligent behavior can be generated in a system, by giving it a symbolic representation of its environment and its desires, and allowing it to syntactically manipulate this representation. The symbolic AI paradigm rests upon the physical-symbol system hypothesis, formulated by Newell and Simon.

A deductive or deliberative agent is one that contains an explicitly represented, symbolic model of the world, in which the decisions are made through logical reasoning, based on pattern matching and symbolic manipulation. In most the cases, these symbolic representation are logical formulae and the syntactic manipulation corresponds to logical deduction or theorem proving. The idea of deliberative agents as theorem provers is attractive and a number of more-or-less "pure" logical approaches to agent programming have been developed. However, there are several problems associated with this approach to agency to be solved (much of them come from the symbolic approach to AI), some of them are:

The transduction problem: how to translate the real world into an accurate, adequate symbolic description, in time for that description to be useful.

The representation/reasoning problem: how to symbolically represent knowledge about complex and dynamic real-world entities and processes, and how to get agents to reason with this knowledge in time.

Calculative rationality: the assumption that the world will not change in any significant way while the agent is making decisions. This is not acceptable in dynamics environments that change faster.

Computational complexity: the complexity of theorem proving makes it questionable whether agents using this deduction mechanism can operate effectively in time-constrained environments.

2. Reactive architectures

The problems with symbolic or logical approaches to building agents led some researches to proclaim that a whole new approach was required. These researchers began to investigate different alternatives to the symbolic AI paradigm.

Although it is difficult to characterize these different approaches, they agree in certain points:

- the rejection of symbolic AI (as representation-reasoning mechanism).
- intelligence, rational behavior is not disembodied (is a product of the interaction the agent maintains with its environment).
- intelligent behavior emerges from the interaction of various simpler behaviors.

Because these alternative systems are often perceive as simply reacting to the environment, without reasoning about it, these approaches are sometimes called *reactive architectures*. One of the best-known reactive agent architecture is the *subsumption architecture*, developed by Rodney Brooks [10], one of the most influential critics of the symbolic approach to agency in the last years. There are two defining characteristics of the subsumption architecture. The first is that an agent's decision making is realized through a set of *task accomplishing behaviors*. Each behavior may be though of as an individual action function, takes perceptual input and maps it to an action, neither does it includes any complex symbolic representation nor reasoning. Each of these behavior modules is intended to achieve some particular task. In Brook's implementation these modules are finite state machines.

The second important characteristic is that many behaviors can fire simultaneously. Hence, there must be a control mechanism to choose between the different actions selected. Brooks proposed arranging the modules into a *subsumption hierarchy*, with the behaviors arrange into *layers* —the lower the layer is, the higher is its priority. Another characteristic of the subsumption systems implementation is that there is assumed to be a quite tight coupling between perception and action, and there is no attempt to transform the input data to symbolic representations.

One of the principal advantage of these approaches over the logic-based ones is that the complexity is tractable. The overall time complexity of the subsumption action function is no worse than $O(n^2)$ where $n = max(n_b, n_p)$, n_b is the number of behaviors and n_p is the number of percepts. Other advantages of the reactive approaches such that Brooks' subsumption architectures are: simplicity, economy and robustness against failure; making this kind of architectures attractive. However, there are some fundamentals unsolved problems related to the reactive architectures that are remarked in [76]:

- If reactive agents do not employs models of their environment, then they must have sufficient information available to determine an acceptable action
- How reactive agents would take into account non-local information as these agents make decision based on local information.
- How to incorporate learning from experience
- It is difficult to engineer this kind of agents to fulfill specific tasks and there is no principled methodology for building reactive agents. In purely reactive systems the overall behavior emerges from the interaction of component behaviors when the agent is placed in its environment. Sometimes the relationships between individual behaviors, environment and the overall behavior is not understandable.
- It is hard to build agents that contains many layers. Effective agents can be generated with small –less than ten– numbers of behaviors.

This area of work is out of the mainstream of traditional AI work and is documented primarily in the artificial life literature.

3. Layered Architectures - Hybrid agents

Many researches have argued that neither a complete reactive nor deliberative approach is suitable for building agents. Given the requirement that an agent must be capable of reactive and proactive behavior, an interesting approach involves creating separate subsystems to deal with these different kinds of behaviors. A class of architectures in which the defined subsystems are arranged into hierarchy and interacting *layers*, implements this idea.

In this approach, an agent will be defined in terms of two or more layers, to deal with the reactive and pro-active behaviors, respectively. The agent's control subsystems are arranged into a hierarchy, with higher layers dealing with information at increasing levels of abstraction. An important problem in such architectures is to determine what kind of control framework is needed, in order to manage the interactions between the various layers. Two basic types of control flow can be identify within layered architectures, as it is shown in Figure 2.1 and are described in [76]:

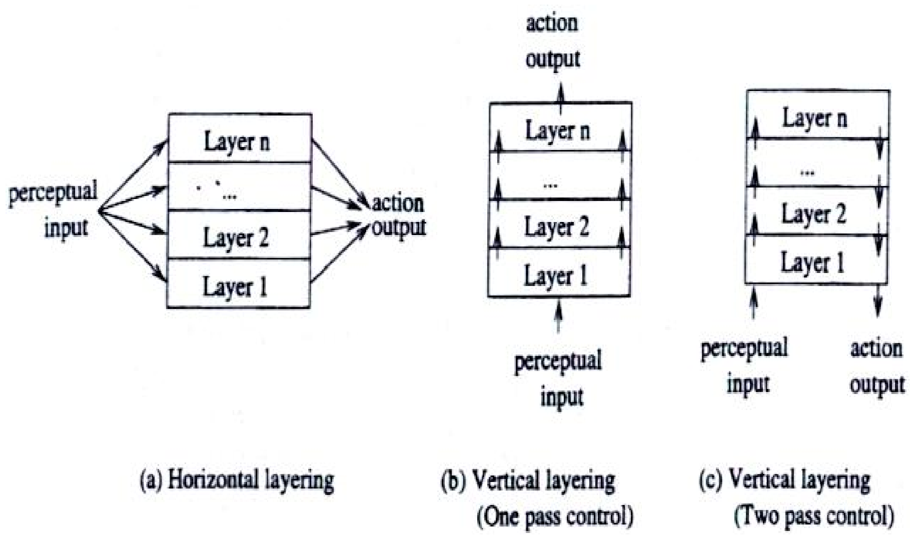

Figure 2.1: Information and control flows in layered agents architecture.(Source: [54], p263).

- *Horizontal layering.* Layers are each directly connected to the sensory input and action output. Each layer itself acts like an agent, producing actions proposals (Figure 2.1(a)).
- *Vertical layering.* Sensory input and action output are each dealt with by at most one layer. In this case there are two approaches:
  - one-pass architecture: control flows sequentially through each layer, until the final layer (Figure 2.1 (b))
  - two-pass architecture: control flows up the architecture (the first pass) and then, control flows back down (Figure 2.1 (c))

The great advantage of the horizontally layered architecture is it conceptually simplicity. One layer can be implemented for each behavior the agent needs to exhibit. The different layers may generate competitive actions suggestions, sometimes inconsistent. In order to ensure the system consistence, it generally includes a mediator function. This function decides which layer has control on the agent at any time, the definition of this function is difficult to design. This problem is in part solution in the vertically layered architecture where the complexity of interactions between layers is reduced. However, the vertical layering has a disadvantage: in this architecture the control must pass between each different layer and a failure in any one layer will be transfer to the agent performance. Examples of the layered architectures are Ferguson's Tourng-Machines —horizontally layered architecture— [24], and Muller's InteRRaP —two-pass vertically layered— [55].

Practical Reasoning Architectures (BDI agents)

There is a particular model of decision making knowing as practical reasoning. These model is inspired in the process that seems to take place as we decide what to do —the process of deciding, time to time, which action to perform in order to reach our goals. The philosopher Michael Bratman defines this process as:

> 'Practical reasoning is a matter of weighing conflicting considerations for and against competing options, where the relevant considerations are provided by what the agent desires/values/care about and what the agent believes.' (Bratman, [9], p.17)

Practical reasoning involves two important processes: *deliberation* —deciding what goals we want to achieve, and *means-ends reasoning* —how we are going to achieve these goals. After generating these set of alternative goals, the agent must choose between them, and commit to some. These committed to achieve goals are the agent's intentions. In practical reasoning process, intentions play a crucial role, fundamentally they tend to lead to actions.

Belief-Desire-Intention (BDI) architectures were originated in the work of the Rational Agency Project at Stanford Research Institute in the mid-1980s. The origins of the model are due to the theory of human practical reasoning developed by M. Bratman [7, 9], which focuses particularly on the role of intentions. Specifically, Bratman argued that rational agents will tend to focus their practical reasoning on the intentions they have already adopted, and will tend to avoid the consideration of options that conflict with them. Some of the most relevant points of his work are knowing as Bratman's claim, and we can summarize their characteristics following [76]:

- *Intentions drive means-ends reasoning.* If an agent has an intention, then it will attempt to achieve it, which involves deciding how to achieve it. If one way fails to achieve an intention, then it will attempt others.

- *Intentions persist.* The agent will not give up on its intention without a good reason –it believes it cannot achieve them or that the reason for the intention is no longer present.

- *Intentions constrain future deliberation.* The agent will not consider options that are inconsistent with its current intention, and

- *Intentions influence beliefs upon which future practical reasoning is based.* The agent can plan for the future on the assumption that it will achieve the intention.

IRMA (for the "Intelligent Resource-Bounded Machine Architecture described in [7]) is an specific BDI agent architecture that embody Bratmans claim.

In the Agent Community, the term BDI model is used in a narrow sense —as in the IRMA specific architecture, and in a wide sense. In this report, we adopt the last position, calling Belief-Desire-Intention (BDI) models, to the models of practical reasoning that employ the folk-psychology concepts of belief, desire and intention, perhaps among other attitudes.

Then, the basic components of a BDI architecture are data structures representing the beliefs, desires and intentions of the agent, and functions that represent its deliberation –deciding what intentions to have, and means ends reasoning –deciding how to do them. In the following we described a general formalization of this components and its relations. Let Bel be the set of all possible beliefs, Des be the set of all possible desires, Int be the set of all possible intentions and P is the current percept. The state of a BDI agent at any given point of time is a triple (B, D, I), where $B \subseteq Bel$, $D \subseteq Des$ and $I \subseteq Int$. The process of practical reasoning in a BDI agent may be summarized as in [72] in the schema shown in Figure 2.2. This Figure illustrates the main components in a BDI agent that are described as follows:

- a *set of current beliefs (B)*, representing the information the agent has about its environment;

- a *belief revision function (brf)*, which takes a perceptual input and the agent's current beliefs, and determines a new set of beliefs:
  $brf : \wp(Bel) \times P \to \wp(Bel)$;

- *an option generation function (options)*, which determines the options available to the agent (its desires–D), on the basis of the current beliefs

and its current intentions:
$options : \wp(Bel) \times \wp(Int) \rightarrow \wp(Des)$

- a *set of current options*, representing possible course of actions available to the agent;
- a *filter function (filter)*, which represents the agent's deliberation process in which the agent determines its intentions, based on its currents beliefs, desires and intentions:
  $filter : \wp(Bel) \times \wp(Des) \times \wp(Int) \rightarrow \wp(Int)$
- a set of *current intentions (I)*, representing those state of affairs that it has committed to trying to bring about.
- an *action selection function (execute)*, which determines an action to perform on the basis of current intentions.
  $execute : \wp(Int) \rightarrow A$

The resulting process is the agent decision function $action : P \rightarrow A$. This function maps the input perception into an action that the agent will try to execute and is defined in terms of the data structures and functions previously presented. A simple version of this function is defined by the following pseudocode:

```
function ACTION (p:P):A
    B:= brf(B,p)
    D:= options(D,I)
    I:= filter(B,D,I)
    return execute(I)
end function ACTION
```

A more complete version of this practical reasoning loop, including intention reconsideration, could be seen in ([76], p76).

In the following Section we describe in some detail the formalization of the BDI model proposed by Rao and Georgeff and we show its logical framework.

2.1.3 Rao and Georgeff's BDI model

In the design of rational agents the role played by attitudes such as beliefs(B), desires (D) and intentions (I) has been well recognized and analyzed by philosophical and AI researchers. The beliefs are needed to represent the state of the

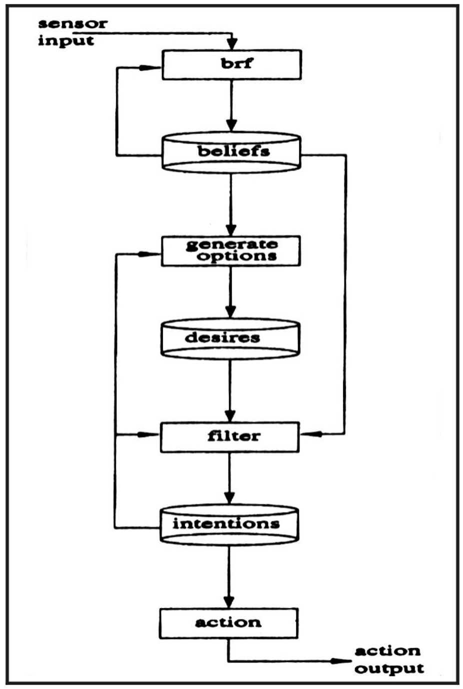

Figure 2.2: Schematic diagram of a generic belief-desire-intention architecture.(Source: [72], p58).

world, the desires, to set the state of affairs the agent wants to achieve and the intentions, the worlds the agent has chosen and is committed to achieve.

In these direction Bratmans theory of intention [7, 9] is an important contribution. Its work points out that intentions play an fundamental role in practical reasoning. As it can be seen (for instance, in the *option* and *filter* functions defined in page 12) the intentions interact with an agents beliefs and desires. However, satisfactorily capturing these interactions turn out to be considerably difficult.

Some of the philosophical aspects of the theory were well formalized by Cohen and Levesque [17, 18]. They construct a *logic of rational agency* [17] where they provided one of the first logical formalization of intentions and the notion of commitment using just two basic attitudes: beliefs and goals (i.e., desires). Other attitudes, as commitments or intentions, were defined in terms of these. In particular, intentions are modeled as a kind of commitment (i.e., persistent goal) and are defined in terms of temporal sequences of an agents beliefs and goals. Based on different kinds of commitments they defined specific agents. An agent *fanatically committed* to its intention will maintain its goal until either they are believed to be achieve or are believed to be unachievable; an agent with a *relativized commitment* to its intentions is similarly than the other but may also drop its intentions when some specified conditions are believed to hold. This theory of intention and commitment was applied in the formalization of communicative actions for agents [18].

While this formalization treat intentions as being reducible to beliefs and desires, Bratman [7] argues that intentions plays a significant and distinct role in practical reasoning. He also shows how the agents existing beliefs, desires and intentions form a background for future deliberations. Systems and formalisms that give primary importance to intentions represent an important class of the BDI architectures. One of the well-known intentional system formal approach, that follows Bratman's claim, was proposed by Rao and Georgeff [61, 63]. Below, we outline the more important features of this formalism presented in [61].

Formal framework for BDI agents

This model is based on the explicit representation of the agent's beliefs (B), desires (D) and intentions (I), using a logical framework based on the possible world semantics.

The traditional possible-worlds semantics of beliefs, considers each world to be a collection of propositions and models belief by a belief-accessibility relation B linking these worlds. Cohen and Levesque [17] treat each possible world as a time-line representing a sequence of events, temporally extended infinitely into the past and the future.

Rao and Georgeff used to model the world, a temporal structure with a branching time future and a single past, called a time tree. The branches in a time tree can be viewed as representing the choices available to the agent at each moment of time. Event types transform one time point into another. Primitive events (actions) are those events that the agent can execute directly, and uniquely determine the next time point in a time tree. Non-primitive events (plans) map to non-adjacent time points. The agent may attempt to execute some event, but may fail to do so (i.e., successful execution of events or their failure). They use a formalism similar to Computation Tree Logic, CTL [23] to describe these structures. A distinction is made between state formulas — evaluated at a specified time point in a time tree, and path formulas —over a specified path in a time tree.

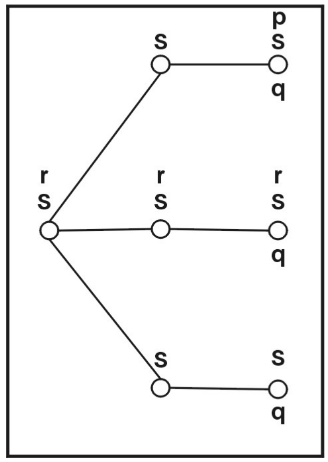

Figure 2.3: Belief-accessible world (Source: [61]).

The modal operators *optional* and *inevitable* are introduced to operate on path formulas. A path formula ψ is said to be optional if, at a particular time point in a time tree, ψ is true in at least one path emanating from that point; it is inevitable if ψ is true in all paths emanating from that point. They also introduced the standard temporal operators O(next),$\Diamond$(eventually), $\Box$(always) and U(until), in order to operate over state and path formulas.

These modalities can be combined in various ways to describe the options available to the agent. For example, as is showed in [61], the structure illustrated in the Figure 2.3 could be used to represent the following statements:
'it is optional that John will eventually visit London (denoted by p); it is optional that Mary will always live in Australia (r); it is inevitable that the world will eventually come to an end (q); and it is inevitable that one plus one will always be two (s).'

Belief is modeled in the conventional way, in each situation they associate a set of belief-accessible worlds and each belief-accessible world is a time tree. Multiple belief-accessible worlds result from the agent's lack of knowledge about the state of the world. They take in account the uncertainty in the agent's beliefs allowing a set of possibles worlds, but in this approach they do not use a belief measure (e.g., a probability measure, possibility measure, etc) to establish an order over the set of worlds (i.e., expressing which of these are the most believable ones). Within each of these worlds, the branching future represents

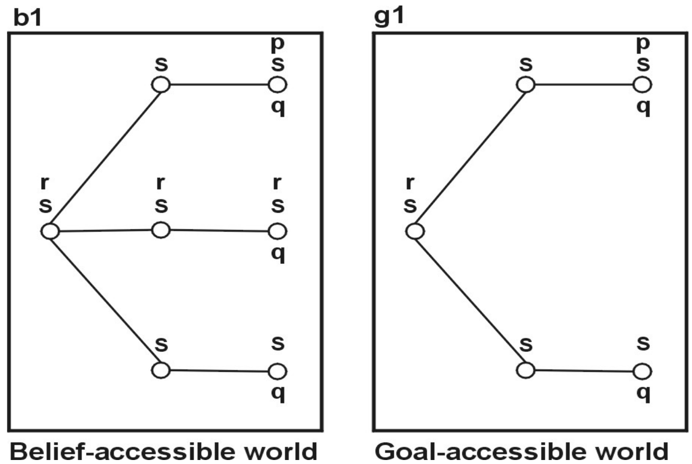

Figure 2.4: Compatibility between Belief and Goal-accessible world (Source: [61]).

the choice (options) still available to the agent in selecting which actions to perform. In similar way, for each situation they associate a set of goal-accessible worlds to represent the goals of the agent. They use goals as a set of chosen consistent desires. In this review, we adopt the notion of *strong realism*. This sets up a relation between the belief- and goal-accessible worlds: it is required that the agent believes it can optionally achieve its goals. This kind of belief-goal compatibility is illustrated in Figure 2.4

Intentions are similarly represented by sets of intention-accessible worlds. These worlds are ones that the agent has committed to attempt to realize. The intention-accessible worlds of the agent must be compatible with its goal-accessible worlds. In the case of a *strong realism* agent, it can only intend some course of action if it is one of its goals.

It thus remains to be shown how these attitudes determine the actions of an agent and how they are formed, maintained, and revised as the agent interacts with its environment.

BDI logic

The CTL logic is extended in two ways. First, they consider a first-order variant of the logic, and second, it is extended to a possible-worlds framework by introducing modalities for the believes, desires and intentions. There are two types of formulae in the logic: *state* formulae and *path* formulae. A *state*formula is defined as follows:

- any first-order formula is a state formula,
- if φ and ψ are state formulas and x is an individual or event variable, then $\neg\varphi$, $\varphi \vee \psi$, and $(\exists x)\psi(x)$ are state formulas,
- if e is an event type then succeeds(e), fails(e), does(e), succeeded(e), failed(e), and done(e) are state formulas;
- if ψ is state formula then $BEL(\psi)$, $GOAL(\psi)$ and $INTEND(\psi)$ are state formulas; and
- if ψ is a path formula, then $optional(\psi)$ is a state formula.

A path formula can be defined as follows:

- any state formula is also a path formula, and
- if Φ and Ψ are path formulas, then $\neg\Phi$, $\Phi \vee \Psi$, Φ U Ψ, $\Diamond\Psi$ and OΨ are path formulas.

Possible-worlds Semantics

The formalization of this semantics is presented by Rao and Georgeff in [61]. First, they provide the semantics of the different formulae, secondly of the events and finally, the possible-worlds semantics of beliefs, goals, and intentions. In the following we briefly outline this schema:

An interpretation M is defined as $M = (W, E, T, \prec, U, B, G, I, \Phi)$, where:

- W is a set of worlds,
- E is a set of primitive event types,
- T is a set of time points,
- $\prec$ a binary relation on time points,
- U is the universe of discourse,
- Φ is a mapping of first-order entities to elements in U for any given world and time point, and

- $B, G, I \subseteq W \times T \times W$ are accessible relations for BEL, GOAL and INTEND, respectively.
  Notation: R refers to any one of these relations and R_t^w to denote the set of worlds R-accessible from world w at time t.

Each world $w \in W$, called a time tree, is a tuple (T_w, A_w, S_w, F_w), where:

- $T_w \subseteq T$ is a set of time points in the world w,
- A_w is the same as $\prec$ restricted to T_w,
- $S_w : T_w \times T_w \rightarrow E$ map adjacent time points to (successful) events in E and
- $F_w : T_w \times T_w \rightarrow E$ map adjacent time points to (failing) events in E.
  The domains of S_w and F_w are disjoint.

Considering an interpretation M and a variable assignment v, the semantics of the state formulae are defined as following:

- $M, v, w_t \models q(y1, ..., yn) \Leftrightarrow (v(y_1), ..., v(y_n)) \in \Phi[q, w, t]$ where $q(y1, ..., yn)$ is a predicate formula.
- $M, v, w_t \models \neg\phi \Leftrightarrow M, v, w_t \neg \models \phi$
- $M, v, w_t \models \phi \vee \psi \Leftrightarrow M, v, w_t \models \phi$ or $M, v, w_t \models \psi$
- $M, v, w_t \models (\exists x)\phi \Leftrightarrow M, v\,[d/x]\,, w_t \models \phi$ for some $d \in U$
- $M, v, w_t \models optional(\phi) \Leftrightarrow$ exists a full path $(w_{t0}, w_{t1}, ...)$ such that $M, v, (w_{t0}, w_{t1}, ...) \models \phi$

Semantics of events:

- $M, v, w_{t1} \models succeeded(e) \Leftrightarrow$ exists $t0$ s.t. $S_w(t0, t1) = e$
- $M, v, w_{t1} \models failed(e) \Leftrightarrow$ exists $t0$ s.t. $F_w(t0, t1) = e$

Semantics of Belief, Goals and Intentions:

The possible-worlds semantics of beliefs, considers each world to be a collection of propositions and models belief by a belief-accessibility relation B linking these worlds. In this BDI model, each possible world is a time tree and denotes the optional courses of events that an agent can choose in a particular world. The belief relation maps a possible world at a time point to other possible worlds. An agent has a belief ϕ, denoted $BEL(\phi)$, at time point t if and only if ϕ is true in all the belief-accessible worlds of the agent at time t. The semantics of the modal operator $GOAL$ is given in terms of a goal-accessible relation G which is similar to that of the B relation. The goal-accessibility relation specifies situations that the agent desires to be in. Intentions can be seen as future paths that the agent chooses to follow. The intention-accessibility relation will be used to map the agent's current situation to all its intention-accessible worlds. Formally, this semantics is defined as follows:

- $M, v, w_t \models BEL(\phi) \Leftrightarrow \forall w' \in B_t^w,\ M, v, w'_t \models \phi$
- $M, v, w_t \models GOAL(\phi) \Leftrightarrow \forall w' \in G_t^w,\ M, v, w'_t \models \phi$
- $M, v, w_t \models INTEND(\phi) \Leftrightarrow \forall w' \in I_t^w,\ M, v, w'_t \models \phi$

The semantics of path formulae:

- $M, v, (w_{t0}, w_{t1}, ...) \models \phi \Leftrightarrow M, v, w_{t0} \models \phi$ (ϕ state formula)
- $M, v, (w_{t0}, w_{t1}, ...) \models O\phi \Leftrightarrow M, v, (w_{t1}, w_{t2}, ...)\phi$
- $M, v, (w_{t0}, w_{t1}, ...) \models \Diamond\phi \Leftrightarrow M, v, (w_{tk}, ...) \models \phi$ for some $k \geq 0$
- $M, v, (w_{t0}, w_{t1}, ...) \models \phi$ U $\psi \Leftrightarrow$
  - exists $k \geq 0$ s.t. $M, v, (w_{tk}, ...) \models \psi$ and $\forall 0 \leq j \leq k$ $M, v, (w_{tj}, ...) \models \phi$, or
  - $\forall j \leq 0$, $M, v, (w_{tj}, ...) \models \phi$

Axiomatization

The basic axiomatization for beliefs is the classic weak-S5 modal systems or KD45. For goals and intentions are adopted the K and D axioms, to make them closed under implication and to satisfy the consistence condition. There is also needed the rule of necessitation for beliefs, goals and intentions (i.e., the agent believes, has as goal and intends all the valid formulae). As happened in most of possible worlds formalisms, this logic suffers from the logical omniscience problem (i.e., the agent believes, desires and intends all the logical consequences of its beliefs, desires and intentions). There are several alternatives to solution this problem, but they are not presented in this review. Then, the axiom schema is the following:

- $BEL(\phi \rightarrow \psi) \rightarrow (BEL\phi \rightarrow BEL\psi)$ (axiom K)
- $BEL\phi \rightarrow \neg BEL(\neg\phi)$ (consistence - axiom D)
- $BEL\phi \rightarrow BEL(BEL\phi)$ (positive introspection - axiom 4)
- $\neg BEL\phi \rightarrow BEL(\neg BEL\phi)$ (negative introspection - axiom 5)
- $GOAL(\phi \rightarrow \psi) \rightarrow (GOAL\phi \rightarrow GOAL\psi)$
- $GOAL\phi \rightarrow \neg GOAL(\neg\phi)$
- $INTEND(\phi \rightarrow \psi) \rightarrow (INTEND\phi \rightarrow INTEND\psi)$
- $INTEND\phi \rightarrow \neg INTEND(\neg\phi)$
- Necessitation rule for beliefs, goals and intentions (from ϕ derive $BEL\phi$, $GOAL\phi$ and $INTEND\phi$)

Rao and Georgeff in [61] presented a set of axioms (A11 and A12) in order to set the interrelations among an agent's belief, goals and intentions. They also added an axiom that leads intentions to actions (A13), and two axioms (A14-A15) where is established that the agent believes what it is intending and which are its goals. This group of axioms is:

- (A11) $GOAL(\alpha) \rightarrow BEL(\alpha)$ (belief-goal compatibility)
- (A12) $INTEND(\alpha) \rightarrow GOAL(\alpha)$ (goal-intention compatibility)
- (A13) $INTEND(does(e)) \rightarrow does(e)$ (intention leading to action)
- (A14) $INTEND(\phi) \rightarrow BEL(INTEND(\phi))$
- (A15) $GOAL(\phi) \rightarrow BEL(GOAL(\phi))$
- (A16) $INTEND(\phi) \rightarrow GOAL(INTEND(\phi))$
- (A17) $done(e) \rightarrow BEL(done(e))$ (awareness of primitive events)

- (A18) $INTEND(\phi) \rightarrow inevitable\Diamond(\neg INTEND(\phi))$ (no infinite deferral)

This set of eight axioms A11-A18 together with the standard axioms for BDI logics (KD45 for BEL and K-D for GOAL and INTEND) constitute the basic I-system. Furthermore, Rao and Georgeff analyzed in [61] the relation between current and future intentions —commitment strategy— in a process of intention maintenance and revision. They described three different commitment strategies: *blind*, *single minded* and *open minded*. A *blindly committed* agent is one who maintains it's intentions until it actually believes that it has achieved them. A *single-minded commitment*, is an agent which maintains it's intentions as long as it believes that they are still options. Finally, an *open-minded* agent is one who maintains it's intentions as long as these intentions are still it's goals. In order to obtain one of these different behaviors in an agent it must be added the corresponding axiom to the basic I-system:

- $INTEND(inevitable\Diamond\phi) \rightarrow$
$inevitable(INTEND(inevitable\Diamond\phi)$ U $BEL(\phi))$.
(for a Blind agent)

- $INTEND(inevitable\Diamond\phi) \rightarrow$
$inevitable(INTEND(inevitable\Diamond\phi))$U $(BEL(\phi) \vee \neg BEL(optional\Diamond\phi))$.
(for a single-minded agent)

- $INTEND(inevitable\Diamond\phi) \rightarrow$
$inevitable(INTEND(inevitable\Diamond\phi))U(BEL(\phi) \;\vee\; \neg GOAL(optional\Diamond\phi))$.
(for an open-minded agent)

Advantages of BDI models

Several factors have contributed to the importance of the BDI model. This architecture is one of the best models of practical reasoning that is based on well understood logical foundations. The BDI model has proved to have the essential components to cope with the complex, real world applications. These real systems are usually placed in a dynamic and uncertain environment, having a local view of the world and are resource bounded. These constrains have certain fundamental implications for the design of the underlying computational architecture, and the Belief's, Desire's and Intention's components seem to be an essential part of such systems.

The BDI model is also interesting because a great deal of effort has been done in its formalization. In particular, Rao and Georgeff have developed a range of BDI logics. They set out different axiomatics to set up properties of the BDI agents (e.g., different commitment strategies).

But the importance of the BDI models is not limited to the theoretical field. Since the end of 80's there have been different developments of particulars BDI architectures. One of the specific BDI agent architecture is IRMA, described by Bratman et al. in [8], this architecture has been evaluated in an experimental scenario known as the Tileword. However, the best-known implementation of the BDI model is the Procedural Reasoning System (PRS) system developed by M. Georgeff and A. Lansky [25]. The PRS system has been re-implemented several times afterwards, as for example in dMARS system [21] and in the java version called Jam system [38]. This agent architecture has proved to be the most durable agent architecture developed to date. It has been applied in several of the most significant multiagent applications built up to now, including OASIS, an air-traffic management system [48], and SPOC (Single Point of Contact), a business process management system [26].

The BDI architecture has evolved over time and diverse factors have contributed to the importance of this model as it is summarized by Georgeff et al. in the Fifth International Workshop on Agent Theories, Architectures, and Languages ATAL-98 [27]:

> The belief-desire-intention (BDI) model has come to be possibly the best known and best studied model of practical reasoning agents. There are several reasons for its success, but perhaps the most compelling are that the BDI model combines a respectable philosophical model of human practical reasoning, (originally developed by Michael Bratman [7]), a number of implementations (in the IRMA architecture [8] and the various PRS-like systems currently available [25]), several successful applications (including the now-famous fault diagnosis system for the space shuttle, as well as factory process control systems and business process management [26]), and finally, an elegant abstract logical semantics, which have been taken up and elaborated upon widely within the agent research community [64, 68].

Because the recognized relevance of the BDI model it was decided to use this agent architecture to be extended in this research work. In the following subsection we introduce one interesting approach to specify agent's architectures and in particular, BDI agents.

2.1.4 Multi-context Systems

The notion of context has been studied in many research areas an in particularly in Artificial Intelligence. Contexts are view as an important approach to represent certain kinds of reasoning. On the one hand, context are a tool for formalizing the locality of reasoning. While on the other hand, context are introduced as a mean of solving the problem on generality. Coherently with these two points of view, Fausto Giunchiglia et al. in [30, 31] introduced the notion of *multi-context system* (MCS for short). These systems have also been called multi-language systems in [31], in order to emphasized that they may include multiple languages. Contexts have been used in diverse applications as: integration of knowledge and data bases and in the formalization of reasoning about beliefs, among others. Particularly, have been used to model different aspects of agents and multiagent systems [16, 2]. Despite the different approaches, formalizations and applications, there are two main intuitions underlying the use of contexts, called principles in [28]:

- *Locality principle*: reasoning uses only part of what is potentially available (e.g.,what is known, the available inference procedures). The part being used while reasoning, is what we call context (of reasoning);
- *Compatibility principle*: there is a compatibility among the kinds of reasoning performed in the different contexts.

These two principles are formalized by the semantics called *Local Models Semantics*, which is described in [28]. In this paper the authors also showed how this novel semantics is capture by the MCS. They also validates this semantics by formalizing two important forms of contextual reasoning: reasoning with viewpoints ans reasoning about belief.

One of the advantage of the MCS in order to help in the design of complex logical systems is, that this framework allows the definition of separately and different formal components, and their corresponding interrelations are neatly specified.

Formalization of multi-context systems

We present an introduction to the formal aspects of MCS systems, where contexts are formalized proof-theoretically. A more complete description is given in [31]. The MCS specification contains three basic components: units or contexts, logics, and bridge rules, which channel the propagation of consequences among theories. Following this, a MCS is defined as a group of interconnected units:
$\left\langle \{C_i\}_{i \in I}, \Delta_{br} \right\rangle$, where:

- for each $i \in I$, $C_i = \langle L_i, A_i, \Delta_i \rangle$ is an axiomatic formal system where L_i, A_i and Δ_i are the language, axioms, and inference rules respectively. They define the logic for the context C_i and its basic behavior is constrained by the axioms.

- Δ_{br} is a set of bridge rules, they are rules of inference which relates formulae in different units.

When a theory $T_i \in L_i$ is associated with each unit, the specification of a particular multi-context system is complete. A MCS system is essentially a set of logical theories, plus a set of inference rules which allow for the propagation of consequences among theories.

The *bridge rules* 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 machinery Δ in these systems is then based on two kinds of inference rules, internal rules Δ_i inside each unit, and bridge rules Δ_{br} outside, i.e.,

$$\Delta = \bigcup_{i \in I} \Delta_i \cup \Delta_{br}$$

Internal rules allow to draw consequences within a theory, while bridge rules allow to embed results from a theory into another [28].

A MCS formalized the principle of locality in the sense that each context has it suitable language L_i, the proper set of axioms A_i which provides the foundations of reasoning, the theory $T_i \subset L_i$ setting the true formulae in each context, and finally, the inference engine Δ_i, capturing different deduction capabilities for each unit. Through bridge rules is represented the principle of compatibility, these rules allow contexts mutually influence themselves. Bridge rules change the set of formulae in one context by the derivation of formulae in other contexts.

MCS have been used in diverse applications as for example in the integration of heterogeneous knowledge and data bases, in the formalization of reasoning about beliefs, among others. Particularly, have been used to model different aspects of agents and multiagent systems [16] and as an approach to engineering multiagent systems [66].

Generic multi-context agent

Multiagent systems are complex systems than can be well modeled by MCS, in order to respect the locality of its architecture components and representing also the interaction betweem them. This approach has been used by Sabater et al. [66] and Parsons et al. [60] to specify several agent architectures and particularly to model some classes of BDI agents [57]. Using the multi-context approach, an agent architecture consist of the following four basic types of components. These components were first identify in the context of theorem provers for modal logic in [31], and in [56] a full detail of these components can be found. In brief, the components are the following:

- *units or contexts*: structural entities representing the main components of the architecture.
- *logics:* declarative languages, each with a set of axioms and rules of inference. Each unit has a single logic associated with it.
- *theories:* subsets of formulae of each language.
- *bridge rules:* rules of inference which relates formulae in different units.

The set of formulae that a given context may contain depends on its initial theory, axioms, inference rules that allow inner deductions; and bridge rules. The formulae introduced by a bridge rule depends on the formulae present in the corresponding contexts in the premise of the bridge rule. Contexts represent the various components of the architecture. They contain the agent's problem solving knowledge, and this knowledge is encoded in the specific theory that the unit encapsulates. In general, the nature of the contexts will vary between architectures.

For example, a BDI agent may have units which represent intentional notions — theories of beliefs, desires and intentions— (as in [57]), whereas an architecture based on a functional separation of concerns may have units which encode theories of cooperation, situation assessment and plan execution (as in [67]). In either case, each context has a suitable logic associated with it.

In any architecture represented, the bridge rules provide the mechanism by which information is transferred between units. The bridge rules continuously examine the theories of the contexts that appear in their premises for new sets of formulae that match them. This means that all the components of the architecture are always ready to react to any change (external or internal) and that there are no central elements of control.

The multi-context approach was used to specified negotiating agents in an example of two Home Improvement Agents, described in [57]. An extended model of multi-context agent was presented in [67] to engineering ReGreT system.

Multi-context BDI agents.

We described in the previous subsection 2.1.3 the BDI architecture. This involved the explicit representation of the agent's beliefs, desires and intentions. In a logical framework this means to include different modalities for the different attitudes and the chosen axiomatic according to the behavior of each attitude. The BDI logic [61] is an example of this, is presented as an unified logical framework and was described in subsection 2.1.3. The axiomatic normally used includes the classic set K, D, 4 and 5 for beliefs; K and D for desires and intentions. As we mention previously, to this basic I-system can be added some axioms to represent the relationships between the attitudes, and some related to commitment strategies. Particularly there are three well-established sets of attitudes relationships for the BDI agents that has been identified in [61] and were presented in a multi-context version in [57]. These three types of agents are illustrated in Figure 2.5 and incorporated different kind of relations between the attitudes, called realisms:

- *Strong realism:* the set of intentions is a subset of desires which in turn the beliefs. That is, if an agent does not belief something, it will neither desire nor intend it.

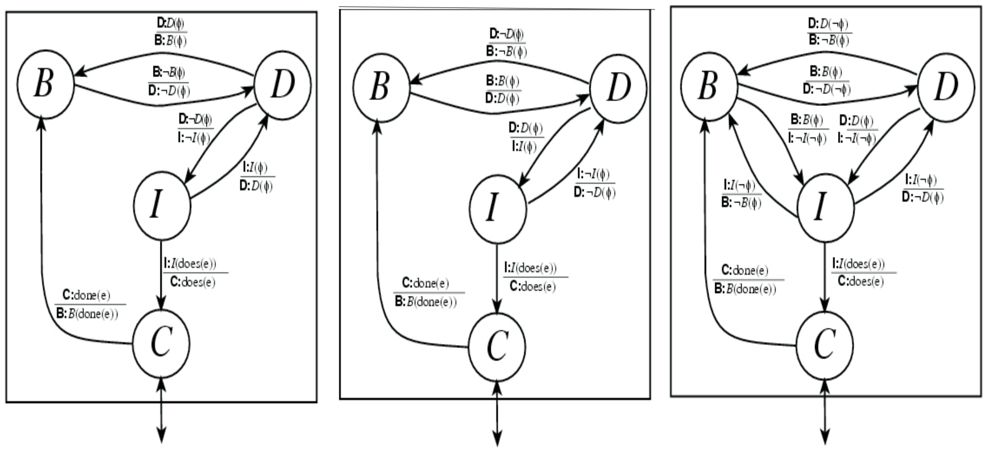

Figure 2.5: Different types of BDI agent. From left to right, the relations between modalities correspond to strong realism, realism and weak realism. (Source [57], p272).

- *Realism:* the set of beliefs is a subset of desires which in turn is a subset of the set of intentions. That is, if an agent beliefs something, it both desires and intends it.
- *Weak realism:* agents do not desires properties if the negation of those properties are believed, do not intend properties if the negation of those properties are desired, and do not intend properties if the negation of those properties are believed.

Modeling different intentional notions by means of several modalities (e.g., B, D and I) can be very complex if only one logical framework (e.g., the BDI logic) is used or if one must manage the pass of formulas of different logics. Using multi-context systems makes it possible to build BDI agents with some advantages over other approaches as it pointed in [66] and exemplify in [57]. This MCS approach enable to use different logics in a way that keeps the logics neatly separated. This either makes it possible to increase the representational power of this BDI agent —compared with those which use a single logic, or simplify agents conceptually —compared with those which use several logics in one global framework.

Thus, in a MCS approach, the belief context of a BDI agent may have a logic of belief associated with it, the desire context may have a logic of preference associated and the same occur for the intention unit. The logic related with each unit provides the language in which the information in that context is encoded.

We illustrate how this approach may be implemented, showing the multi-context BDI agent model presented in [57]. The specification corresponds to a strong realistic BDI agent and its main components can be seen in figure 2.5 and are defined as follows:

Contexts: There are four contexts within a multi-context BDI agent. The units for the beliefs (B), desires (D) and for the intentions (I); and a communication unit (C) is added.

Logics: for each of these four contexts a proper logic is defined:
- *B, D and I context:* each one uses first-order logic with a special predicates B, D and I, which are used to denote respectively the beliefs, desires and intentions of the agent. The chosen axioms are the classics for predicate logics. For capture the behavior of the modalities, in the B context its included the axioms KD45, and in the logics for D and I, the axioms K and D are added.
- *Communication context:* uses classical first-order logic with the usual axioms.

The rules of inference for each unit are the usual ones (MP, MT, generalization, particularization)

Theories: For each context, these logical formulae express the domain information that posses each unit, and depends on the specific agent we are defining (in a generic BDI agent there are not included specific theories).

Bridge rules: The bridge rules are exactly those illustrated in Figure 2.5 for the strong realism BDI agents, formally:

$$I : I(\alpha) \Rightarrow D : D(\lceil \alpha \rceil)$$

$$D : \neg D(\alpha) \Rightarrow I : \neg I(\lceil \alpha \rceil)$$

$$D : D(\alpha) \Rightarrow B : B(\lceil \alpha \rceil)$$

$$B : \neg B(\alpha) \Rightarrow D : \neg D(\lceil \alpha \rceil)$$

$$C : done(e) \Rightarrow B : B(\lceil done(e) \rceil)$$

$$I : I(\lceil does(e) \rceil) \Rightarrow C : does(e)$$

The first four rules derived directly from the model proposed by Rao and Georgeff and ensure compatibility between what the agent believes, desires and intentions. The last two bridge rules specify the interactions that the communication context has with the other units.

Since this generic specification of a BDI agent, concrete agents may be specify as is showed in [57], where the complete specification of two home improvement BDI agents are presented.

In BDI agents, the multi-context approach may it possible to modeled each agent's attitude in an appropriate and local way, and the corresponding interaction between attitudes are neatly represented through the bridge rules. It also allows the incorporation of other attitudes to the agent's model just adding the corresponding contexts and the necessary bridge rules, relating the new attitude with the others.

Advantages of the multi-context specification of agents

Multi-contexts approaches to engineering multiagent systems has some advantages, some of them are pointed by Sabater et al. in [66]. From a software engineering perspective, firstly, MCS support the development of modular architectures. Each architectural component, be it a functional component or a data structure component, can be represented as a separate context. The interrelations between the components can then be made explicit by writing bridge rules to link the contexts. This ability to directly support component decomposition and component interaction offers a path from the high level specification of the architecture to its detailed design (independently on how the architectural components are decomposed or how many components exist). Secondly, MCS are ideally suited to supporting reuse —both of designs and implementations— since these systems encapsulate architectural components and provide specifications for the interrelationships.

From the logical modeling perspective, there are several advantages of adopting a multi-context approach. In first place, separating the logical description of an agent into a set of contexts, each which its proper logic, we effectively get a form of many-sorted logic (all the formulae in one context are a single sort). This brings to the system the advantages of scalability and efficiency. The second advantage comes from the same issue. Using multi-context systems makes it possible to build agents which use several different logics in a way that keeps the logics neatly separated (all the formulae in one logic are gathered in one context). This either makes it possible to increase the representational power of logical agents —compared with those which use a single logic, or simplify agents conceptually —compared with those which use several logics in one global context. This latter advantage was illustrated above, with the description of a multi-context BDI agent.

The remaining two advantages from the logical perspective apply to those logical agents which reason about their mental attitudes and those of other agents. The first is that multi-context systems make it possible to build agents which reason in a way which conforms to the use of modal logics like KD45 (the standard modal logic for handling belief) while working within the computationally simpler framework of standard predicate logic [31]. The final advantage is related to this. Agents which reason about beliefs are often confronted with the problem of modeling the beliefs of other agents, and this can be hard, especially when those other agents reason about beliefs in a different way (because, for instance, they use a different logic). Multi-context systems provide a neat solution to this problem [16, 28].

Combining the software engineering and the logical modeling perspectives, it can be seen that the multi-context approach offers a clear path from specification to implementation.

Indeed one advantage of the MCS logical approach to agency modeling is that allows for rather affordable computational implementation. For instance, a portion of the framework described in [57] has been implemented using a prolog multi-threaded architecture [29].

2.2 Related Works

In this Section, it is reviewed the state of the art of some related works to our model. First, we include the review of some logics approaches that inspired different aspects of our model, such as: the many-valued approach to uncertainty, and a bipolar representation of preferences. Secondly, we present a few works about extensions to the BDI model where some of the attitudes —belief, desires or intentions— where considered to be graded or ordered.

2.2.1 Many-valued approach to uncertainty

In the last two decades, Artificial Intelligence community has undertaken the problem of knowledge representation and reasoning under uncertainty. This was an important and necessary issue, in order to develop systems able to deal with incomplete, uncertainty and vague information in real-domains. There are different approaches to model and manage approximate reasoning. Between the most relevant ones, are the works based on probabilistic models, Dempster-Shafer theory of evidence and possibility theory [50].

Recently, Hájek et al. in e.g. [37] and Gottwald in [32] have developed an approach where uncertainty reasoning is dealt by defining suitable modal theories over suitable many-valued logics. Fuzzy logics and uncertainty theories play different roles that must be clarify.

Fuzzy logic is a logic of vague, imprecise notions and propositions, and is then, a logic of partial degrees of truth. On the contrary, an uncertainty measure dealt with crisp notions and propositions, and it is evaluated with it the degree of belief on the truth of the proposition. Fuzzy logics behave as many-valued logic, whereas uncertainty or belief theories can be related to some kinds of (two-valued) modal logic.

The basic idea presented by Hájek et al. in [37] is the following:

> *Belief degree of a crisp proposition as truth-degree of a fuzzy modal proposition.*

For instance, let us consider the case 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 $B\varphi$ may be seen then as 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.

Therefore, in this kind of logical frameworks we shall have, besides the axioms of the many-valued logic, a set of axioms corresponding to the basic postulates of a particular uncertainty theory.

Following this approach, Hájek et al. in [36] defined a propositional probability logic –Fuzzy Logic of Probability , as a theory over Rational Pavelka logic *RPL* (an extension of Łukasiewic's infinitely-valued logic with rational truth constants). The same authors, in [32] take advantage of a logic combining Łukasiewicz and Product logic connectives –ŁΠ Logic– to define a richer belief theory on top of it, particularly they formalized the logic of conditional probability.

To give an insight of how these logical frameworks are built, we consider some features of the Fuzzy Logic of Probability (FP) presented in [36]. To define the FP-logic language, the authors started with a set of crisp formulae (i.e., built from propositional variables $p, ..., q$ and connectives). They associated with each crisp formula ψ a new propositional variable $P\psi$ (also may be noted f_ψ), meaning "ψ is probable", and which is evaluated using its probability: $e(P\psi) = P(\psi)$. This kind of formulae are named *fuzzy propositional variables* and are taken as the propositional variables of the theory FP.

The syntax of FP-formulae are just *RPL*-formulae built from *fuzzy propositional variables* (i.e., formulae built from variables of the form $P\phi$ using connectives), truth constant $\bar{r}$ for each rational $r \in [0, 1]$ and the use of connectives.

The axiomatic schema presented for this FP-logic is the following:

- (RPL) Axioms of *RPL*. *(Axioms of the many-valued logic)*
- (FP1) $(P\phi, 1)$ for ϕ being an axiom of classical propositional logic.
- (FP2) $(P(\phi \to \psi) \to (P\phi \to P\psi), 1)$ for all ϕ, ψ
  (K axiom for the P modality)
- (FP3) $(P\neg\phi \leftrightarrow \neg(P\phi), 1)$ for ϕ for all ϕ *(Axiom of probability)*
- (FP4) $(P(\phi \vee \psi) \leftrightarrow ((P\phi \to P(\phi \vee \psi)) \to P\psi), 1)$ for all ϕ, ψ
  (Axiom of probability)

Axioms (FP1) and (FP2) guarantee the preservation of classical equivalence and the monotonicity. (FP3) and (FP4) are direct translation of the well-known axiom of probability, the first represents the relationship between the probability of one proposition and its negation, and the second represents the finitely additive property.

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. This many-valued logical framework may be used to represent and reason about degrees in the mental attitudes involved in the agent model, as will be seen later on.

2.2.2 Logics of preference

In Artificial Intelligence the problem of representing and reasoning with preferences has been faced by different researchers [3, 6, 45, 71]. This an important issue when we have to represent the users desires in information systems, or to reason about desires and solve eventually inconsistent goals in multiagent systems.

Preferences may be represented in two forms: positive desires and rejections. Indeed on the one hand, an agent may express what it considers unacceptable (in some degree); and on the other hand, it may express what it considers desire or satisfactory (in some level).

For instance, assume that we want to take a week of holidays and we are looking for a tourist destination in the country. We may provide the tourism agent with two kinds of preferences. In first place, we specify the satisfactory slots, with different levels (we strongly prefer the mountains, moderately prefer the small cities and we weakly like to make rafting), these are positive preferences. Secondly, we describe unacceptable conditions, that are refused in different degrees (as we do not want to travel far than 1000 km). These are negative preferences or rejections.

This bipolar approach has been dealt with by Benferhat et al. in their work about modeling positive and negative information [4]. They presented a framework based on possibility theory where this distinction can be made in a graded way. In logical terms, the two types of information —positive and negative— can be encoded by two types of constraints expressed by necessity measures and other possibility function. Particularly, they applied this model to the representation and fusion of preferences. The description of the bipolar representation of preferences in the possibilistic logic framework can be seen in detail in [3, 5], we briefly outline the relevant features of this approach.

The syntactic specification of these bipolar representation of preferences is done introducing two different sets of equality constraints. These sets corresponds to what the agents rejects and what are its goals or aspirations, respectively:

- $\{R(\phi_i) = \alpha_i, i = 1, ..., n\}$, where ϕ_i is a propositional formula, $\alpha_i \in [0, 1]$ and reflects the rejection strength of ϕ_i by the agent. The higher α_i is, the less acceptable are the solutions satisfying ϕ_i. It turns out that the set of rejections can be handled using the classic possibility and necessity measures.

- $\{G(\psi_j) = \beta_j, j = 1, ..., m\}$, where ψ_j is a propositional formula, $\beta_j \in [0, 1]$ and expresses the minimal level of satisfaction which is guaranteed for the agent, if ψ_j is true. This kind of positive goals cannot be directly handled by the possibilistic logic machinery, they can be represented using the so-called guaranteed possibility function, denoted by Δ.

Modeling rejection in possibilistic logic.

Rejections can be represented, at semantical level, by a total pre-orden on the set of possible outcomes (interpretations). This pre-orden can be encoded in possibility theory using a possibility distribution over the set of interpretations $\pi_R : W \to [0, 1]$. This function π_R associated with a set of rejections **R**= $\{R(\phi_i) = \alpha_i, i = 1, ..., n\}$ is defined:

$\pi_R(w) = 1 - max\{a_i : w \models \phi_i, R(\phi_i) = a_i \in R\}$, with $max\{\emptyset\} = 0$

Clearly, this definition can be viewed in terms of a necessity measure replacing ϕ_i by $\neg\phi_i$ (if $R(\phi_i) = a_i$ then $N(\neg\phi_i) \geq a_i$)

Representing positive goals

The positive goals can also be described in terms of a possibility distribution: $\pi_G : W \to [0, 1]$, $\pi_G(w) \geq \pi_G(w')$ means that w is more satisfactory for the agent than w'. The meaning of $\pi_G(w)$ is different from $\pi_R(w)$, the first evaluates to what degree w is satisfactory for the agent, while $\pi_R(w)$ evaluates to what extend w is acceptable.

The possibility degree π_G associated with a set of positives goals **G**=$\{G(\psi_j) = \beta_j, j = 1, ..., m\}$ is:

$\pi_G(w) = max\{b_j : w \models \psi_j, G(\psi_j) = b_i \in G\}$, with $max\{\emptyset\} = 0$

The addition of positive goals in G can only lead to the increasing of the satisfaction level of w and this is opposite than the behavior of π_R which monotonically decreases with respect to the number of constraints in R.

The set of positive goals cannot be directly handled by standard possibilistic measures. Constraints like $G(\psi_j) = b_i$ are then represented using a function called *guaranteed possibilistic* function, denoted by Δ, first presented by Dubois and Prade in [22] and afterwards, used in [4] to represent bipolar information. The expression $\Delta(\psi) = b$ means that any interpretation where ψ is true, has a satisfaction degree at least equal to b, Then: $\Delta(\psi) = min_{w\models\psi}\pi_G(w)$

Hence, for the disjunction and conjunction Δ behaves as follows :

- $\Delta(\phi \vee \psi) = min(\Delta(\phi), \Delta(\psi))$, so Δ is decreasing with respect to disjunction. Indeed the semantic for disjunctions goes here in an opposite way than in the classical logic. This means that interpretations covered by $\phi\ lor\psi$ are guaranteed to be possible (i.e., because they are observed, feasible, satisfactory, according to the problem) if and only if both the interpretations in ϕ and those in ψ are guaranteed to be possible.

- $\Delta(\phi \wedge \psi) \geq max(\Delta(\phi), \Delta(\psi))$ since the minimum π_G over $\phi \wedge \psi$ may be greater than the minimum over ϕ and over ψ. Applying the maximal specifity principle on Δ, this inequality leads to the equality:
  $\Delta(\phi \wedge \psi) = max(\Delta(\phi), \Delta(\psi))$

This work on bipolarity representation of positives and negatives preferences inspired us to model the agent's positive and negative desires in a multi-context BDI agent, as it be shown in the following chapter.

2.2.3 Graded attitudes in agent's architectures

If we want that agent technologies increase its role in complex and real applications, will have to be considered the difficults there are in real-world environments where the agents interact. Most of them are not completely accessible, non-deterministic and dynamic. Moreover, the preferences or goals of diverse agents (humans or not) interacting in the environment, may be expressed with different level of intensity. This means that there is uncertainty involved not only in the agent's model of the world, but even there are different degrees related to its pro-attitudes. In order to improve the agents performance, we consider important to take into account this uncertainty and graded attitudes in the agent's theory, architecture and implementation.

The agent architectures proposed so far mostly deal with two-valued information. 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. We analyze bellow the relevant features of some of them.

In the BDI model developed by Rao and Georgeff, they explicitly acknowledge that an agent's model of the world is incomplete, by using a branching-time possible-worlds logic, to model the beliefs, goals and intentions. For each situation they associate a set of belief-, goal- and intention-accessible worlds; intuitively, those worlds that the agent believes to be possible, desires to bring about, and commits to achieve, respectively. Multiple possible-worlds results from the agent's lack of knowledge about the state of the world. Within each of these possible world, the branching future represents the choice of action available to the agent.

In a first proposal of the BDI model [61], even they considered the incompleteness of the agent's model of the world, they 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.

afterwards, in [62] the authors extend the expressive power of the BDI logic, introducing subjective probabilities and subjective payoffs to model the process of deliberation. Intuitively, an agent at each situation has a probability distribution on its belief-accessible worlds. The agent then chooses sub-worlds of these belief-accessible worlds that it considers are worth of pursuing, and associates a payoff value with each path. Using a probability distribution on its belief-accessible worlds and the payoff value with each path in its goal-accessible worlds, the agent determines the best plan(s) of action for different scenario. This process is called *Possible-Worlds (PW)* deliberation and is inspired in decision tree theory. The results of this process is a set of the most desirable sub-worlds of the goal-accessible worlds. These sub-worlds are the intention-accessible worlds that the agent commits to achieve. In the cited paper [62] also it is showed the similarity between the PW-deliberation on the one hand, and the decision tree formalism, on the other hand. We consider this is an interesting approach, but has some shortcomings. The first, is that they introduce the probability and payoff in the unified BDI logic framework, increasing its complexity. Secondly, the semantics of the payoff function over the path formulae is not clear. We think the payoff implicitly combines a kind of benefit of achieving some world with the cost of the path. But, as its meaning it is not clear, we consider that may be difficult to determine the function values, and sometimes unnatural. Besides, they no use any measure degrees to represent the intentions, this allowing to obtain an explicitly ordered set of possible intentions results of the deliberation process. And finally, the functions they use in the deliberation process are not neatly related to the BDI model.

Notably, Parsons and Giorgini [58] 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. The paper combines previous authors' works on the use of argumentation in BDI agents, with others approaches to belief revision and updating. The model presented is an extension of the multi-context specification of BDI agents developed in [57], to include degrees of belief. In order to introduce the degrees of belief they translate every proposition in the belief unit (which may contain nested modalities) into an argument with an empty set of grounds. Thus $B(\Phi)$ becomes the argument: $B((\Phi) : : \alpha)$ where α is the associated degree of belief expressed as a mass assignment in Dempster-Shafer theory. They set out the importance of quantifying degrees in desires and intentions, but this is not covered by their work.

Respect to previous works related to graded desires, we mention in first place, the bipolarity representation of preferences due to Benferhat et al. [5]. We found this approach suitable to model agent's desires, and was analyzed in some detail in Section 2.2.2. Lang et al. [46] present another approach to a logical representation of desires, where the notion of hidden uncertainty of desires is introduced. The semantics of this logic is defined by means of two ordering relations representing preference and normality as in Boutilier's logic QDT [6]. 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. It is considered that an ordinal-like uncertainty is present in the notion of plausibility, whose corresponding pre-order may be defined by the proximity of the current world to the set of most plausible (or normal) worlds.

There has been a certain amount of work on the intention reconsideration problem, as for instance in [75], where a formal perspective is presented. afterwards, Parsons et al. in [59] addressed the intention reconsideration in environments which are both complex and dynamic. Other works deal with reasoning about intentions in uncertain domains, as the proposal of Schut et al [70]. They present an efficient intention reconsideration for BDI agents that interact in an uncertainty environment in terms of dynamics, observability, and non-determinism. In this approach they considered that the internal state of an agent consist of beliefs and intentions: $s = \langle Bel, Int \rangle$. The agent's beliefs are represented by a probability distribution $Bel : E \rightarrow [0, 1]$ where E is the set of environment states. The agent's set of intentions Int is a set of environment variables . They assumed that is possible to assign values and cost to the outcomes of intentions: *intention value* $V : Int \rightarrow \Re$ and *intention cost* $C : Int \rightarrow \Re$. They defined the *net value* V_{net} representing the net value of the outcome of an intention i: $V_{net} = V(i) - C(i), i \in Int$. They also express how good is a state defining a *worth* function: $W : S \rightarrow \Re$, the value for each state s is based on the net value of the intentions of the state. Morover, they assumed that one state has an higher worth than another if the net values of all its intentions are higher. From this state representation, they model the intention reconsideration by using the theory of Markov decision process for planning in partially observable stochastic domains (POMDP). They used POMDP approach because the optimality of the policy in this framework, is based on the same three environment's characteristics considered for the intention reconsideration strategy, namely: dynamism, determinism and observability.

All the above mentioned proposals model partial aspects of the uncertainty related to mental notions involved in an agent's architecture.

2.3 Conclusions

We conclude this Chapter 2 with some observations that will help us to outline our research work. With no doubt, the importance of the multiagent technologies have increased in the design and implementation of complex, real systems. In order to achieve the full potential of multiagent approaches, there are some important challenges for the next future. One of them, is to develop agent-based systems using hybrid approaches, strengthening links with other areas of Computer Science, like the uncertainty community in AI; in this direction we placed our research.

After a bibliographic review, we have noticed that there are only a few works dealing with partial aspects of graded attitudes in intentional systems (e.g., uncertainty in beliefs, graded or ordered desires, intention reconsideration in uncertain domains, etc), but we did not find works facing a general model. This has encouraged us to extend an agent architecture to include graded attitudes. Because its aforementioned relevance, we have chosen in first place, to deal with the BDI architecture. In particular, we have opted for a multi-context specification of the BDI model, because this approach shortens the gap between specification and implementation, among other advantages. In the next Chapter 3 we present our proposal of a general model for a graded BDI agent, based in a multi-context specification.

Chapter 3

Exploratory Work

Several previous works have proposed theories and architectures to provide multiagent systems with a formal support. Among them, one of the most widely used is the BDI agent architecture presented by Rao and Georgeff. We consider that an extension of this architecture in order to incorporate degrees in the different attitudes, will not only make the model's semantics richer, but it also will help the agent to take better decisions. With that aim we looked first at the "individual" aspect of agency, and decided to extend the BDI agent architecture to represent and reason under uncertain beliefs and graded motivations. In this Exploratory Work we introduce a general model for graded BDI agents, the consecutive results of this work have been also published in [12, 13, 14, 15]. This model is based on a multi-context specification of agents, and is able to represent graded mental attitudes. In this sense, belief degrees will represent to what extent the agent believes a formula is true. Degrees of positive or negative desire shall allow the agent to set different levels of preference or rejection respectively. Intention degrees shall give also a preference measure but, in this case, modeling the cost/benefit trade off of reaching an agent's goal.

Then, Agents having different kinds of behavior shall be modeled on the basis of the representation and interaction of these three attitudes. The architecture we present will serve as a blueprint to design different kinds of particular agents. In the next Section 3.7, we will illustrate the design process by formalising a simple travel assistant agent.

3.1 Graded BDI agent model

The architecture presented in this research work is inspired by the work of Parsons et.al. [57] about multi-context BDI agents. Multi-context systems and in particular, multi-context specification of BDI agents, were introduced in the subsection 2.1.4 of the previous Chapter.

Following this approach, our BDI model of agent is defined as a group of interconnected units: $\left\langle \{C_i\}_{i \in I}, \Delta_{br} \right\rangle$, where $C_i, i \in I$ are the different contexts and are defined by 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, and Δ_{br} are the bridge rules defining the different inferences through contexts.

In this approach, 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. [37, 32] and described in Section 2.2.1, where uncertainty reasoning is dealt with by defining suitable modal theories over suitable many-valued logics. 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 Łukasiewicz 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 Łukasiewicz 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 in the next sections.

3.2 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 work the particular case of using probability theory as the uncertainty model. Other models might be used as well by just modifying the corresponding axioms.

3.2.1 The BC language

To reason about the credibility of crisp propositions, we define a language for belief representation, following Godo et al.'s [32], based on Łukasiewicz 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 [69] and [53]. 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)

[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, Łukasiewicz 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.

- 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 Łukasiewicz 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 Łukasiewicz logic)

Other Łukasiewicz 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 Łukasiewicz 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).

3.2.2 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 [37].
- $\rho : \Pi_0 \rightarrow 2^{W \times W}$ 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 Łukasiewicz 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 $\|\Phi\|^K = \inf_{w \in W} e(\Phi, w)$.

3.2.3 BC axioms and rules

As mentioned in Section 2.2.1, to set up an adequate axiomatization for our belief context logic we need to combine axioms for the crisp formulae, axioms of Łukasiewicz 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. [35]).
2. Axioms of Łukasiewicz logic for modal formulae: for instance, axioms of Hájek's Basic Logic (BL) [37] plus the axiom: $\neg\neg\Phi \to \Phi$
3. Probabilistic axioms

   $B(\varphi \to \psi) \to_L (B\varphi \to B\psi)$

   $B\varphi \equiv \neg_L B(\varphi \wedge \neg\psi) \to_L B(\varphi \wedge \psi)$

   $\neg_L B\varphi \equiv B\neg\varphi$
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 Łukasiewicz 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. [37]). 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 $\Psi \in T$).

Comparing this axiomatic with the one proposed by Rao and Georgeff for the beliefs in the BDI logic (i.e., KD45 axioms); in our proposal for BC axioms we are also including K and D axioms, but axioms related to introspection (i.e., axioms 4 and 5) are not consider because the BC language does not allow nested modalities.

3.3 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. [5], described in Section 2.2.2, positive and negative information may be modeled in the framework of possibilistic logic. Inspired by this work, we suggest to formalize agent's desires also 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.

3.3.1 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 Łukasiewicz logic as the base logic, but this time extended with a new connective Δ (known as Baaz's connective), considered also in [37]. 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 $\overline{r}$ 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 $\overline{r} \in DC$
  - 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 does not 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.

3.3.2 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. [4]). 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 Łukasiewicz 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 thus $e(D^+\bot, w) = e(D^-\bot, w) = 1$ for all $w \in W$.

3.3.3 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 Łukasiewicz logic axioms extended with Δ for modal formulae. Also, additional axioms characterizing the behavior 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 Łukasiewicz logic with Δ (cf. [37]) for the modal formulae.
3. Axioms for D^+ and D^- over Łukasiewciz logic:
   $D^+(A \vee B) \equiv D^+A \wedge D^+B$
   $D^-(A \vee B) \equiv D^-A \wedge D^-B$
   $\neg_L\Delta(D^+A \wedge D^-A) \to \neg_L(\nabla D^-A \& \nabla D^+A)$, where ∇ is $\neg_L\Delta\neg_L$[2].
   $D^+(\bot)$
   $D^-(\bot)$

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

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 behavior 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. The formalization we present for D^- is somewhat different from the appproach presented by Benferhat et al. in [5], where they used a necessity function (i.e., considering $D^-\phi$ as $N(\neg\phi)$). But in their approach, the second axiom we present in item (3), results from the necessity axiom (i.e., $N(A \wedge B) \equiv N(A) \wedge N(B)$).

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.

3.4 Intention Context

In this context, we represent the agent's intentions. We follow the model introduced by Rao and Georgeff [61, 63], in which an intention is considered a fundamental pro-attitude with an explicit representation. However, as in the work of Cohen and Levesque [17] in our approach, intentions results from the agent's beliefs and desires.

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 toward the goal. A similar semantics for intentions is used in [70], where the net value of an intention is defined as the difference between the value of the intention outcome and the cost of the intention. In [62], this relation is resumed in the payoff function over the different paths. The formalization of the intention's semantics is difficult, because it does not depends only in the formula intended, but also in the plan that the agent execute to achieve a state where the formula is valid. Our work evolved in this aspect as can be seen in [12] and [14].

In our model, 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 $I\psi \to_L I\varphi$ 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.

3.4.1 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 Łukasiewicz 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 toward it.

3.4.2 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 $w \in W$, $\pi_w : W \to [0,1]$ 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 $e : V \times W \to \{0,1\}$ is extended to the non-modal formulae in the usual way. It is extended to modal formulae using Łukasiewicz 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. [37]), that is, the following axioms:

1. Axioms of classical logic for the non-modal formulae.

2. Axioms of Łukasiewicz logic for the modal formulae.

3. Axioms for I over Łukasiewciz logic:
   $I(\varphi \to \psi) \to (I\varphi \to I\psi)$
   $\neg I(\perp)$
   $I(\varphi \wedge \psi) \equiv (I\varphi \wedge I\psi)$

4. Deduction rules are modus ponens and necessitation for I (from φ derive $I\varphi$).

Notice that the K and D axioms proposed to model intentions in the BDI logic, are covered by the axiomatic presented for IC.

3.5 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)* 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 $\varphi \in A$ and c_α is the normalized cost of α.

- *bestplan(φ, α, P, A, c_α)* 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 3.6).

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.

3.6 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 3.1. 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 $\lceil . \rceil$:

$$\frac{B : B\varphi}{P : \lceil B\varphi \rceil} \tag{3.1}$$

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. Furthermore, a filter is used to select the plans with a belief degree of achieving the goal after its execution greater than some b-threshold $-(B([\alpha]\varphi), bthreshold)$. The following bridge rule among D, B, and P contexts does this:

$$\frac{\begin{array}{c} D : \nabla(D^{+}\varphi), D : (D^{-}\psi, \mathit{threshold}), P : \mathit{action}(\alpha, P, A, c), \\ B : (B([\alpha]\varphi), \mathit{bthreshold}), B : B(A \rightarrow \neg\psi) \end{array}}{P : \mathit{plan}(\varphi, \alpha, P, A, c)} \tag{3.2}$$

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:

$$\frac{D : (D^{+}\varphi, d), P : plan(\varphi, \alpha, P, A, c)}{I : (I\varphi, f(d, c))} \tag{3.3}$$

Different functions model different individual behaviors. 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 $f(d, c) = (d + (1 - c))/2$

In BDI agents, bridge rules have been also used to determine the relationship between the mental attitudes and the actual behavior of the agent. Well-established sets of relations for BDI agents have been identified [63]. 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 [61]:

$$\frac{B : \neg B\psi}{D : \neg D\psi} \text{ and } \frac{D : \neg D\psi}{I : \neg I\psi} \tag{3.4}$$

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:

$$\frac{I : (I\varphi, i_{max}), P : bestplan(\varphi, \alpha, P, A, c_{\alpha})}{C : C(does(\alpha))} \tag{3.5}$$

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

$$\frac{C : \beta}{B : B\beta} \tag{3.6}$$

Figure 3.1 shows the graded BDI agent proposed with the different contexts and the bridge rules relating them.

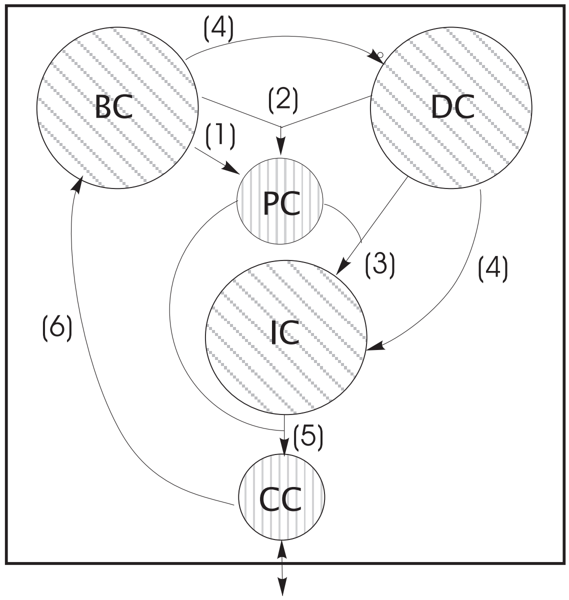

Figure 3.1: Multi-context model of a graded BDI agent

3.7 Example: A Travel Assistant Agent

We design a simple Travel Assistant Agent (T-Agent) using our graded BDI agent model. The T-Agent will be in charge of looking for different holidays plans in Argentinian places, in order to satisfy the desires of different tourists. The plans the T-Agent may offer must be the best choice among the possible plans, taking into account the interests of the tourist, and the cost of each plan.

Suppose a tourist want to instruct the T-Agent to look for a holiday destination package including the transport, that cost less than certain amount of money the tourist is willing to spend, as it is illustrate in the figure 3.2.

In order to obtain the necessary knowledge about the tourist's interests the T-Agent will ask him about his preferences and restrictions, namely:

1- Days available to travel
2- Maximum Cost of the travel
3- Preferences (Regions, Destinations, Activities, etc.)

The tourist can point out if there is a region or even an specific destination

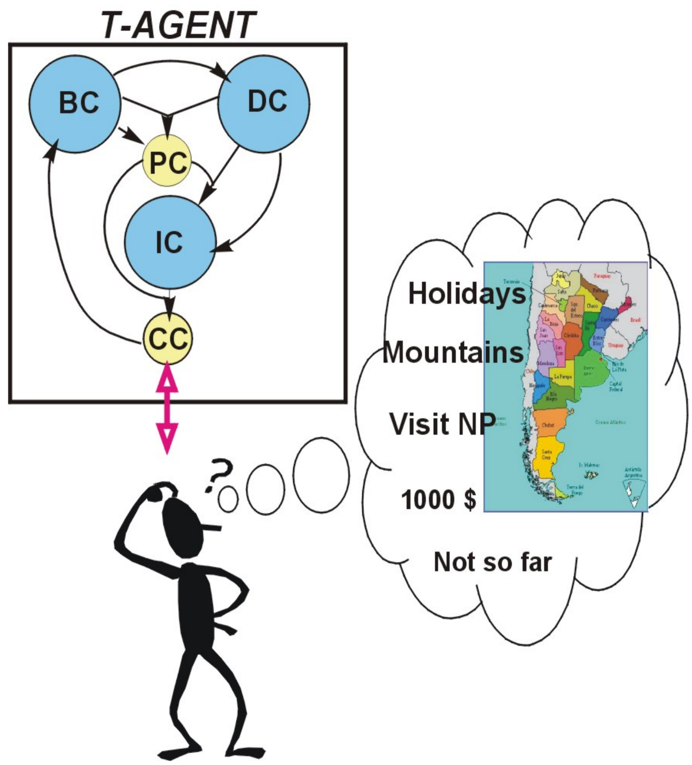

Figure 3.2: A tourist interacting with T-Agent

he prefers or rejects, in some degree to go. These selection will represent some of the positive and negative desires, respectively. If the tourist hasn't got any preferences about the region in the country he may allow the agent to chose the best option without restrictions. For example, the T-Agent may offer the regions and conditions show in the tables 3.1 and 3.2. Each one may be selected or rejected with a degree between 1 to 10.

4- Activities

The tourist is consulted about the desires he has on his holidays. He can introduce degrees (from 1 to 10) in the different answers, the chosen activities will be some of the graded positive desires. The activities offered would vary according to the age of the tourist, the chosen region , the season, etc. For this example we consider the options shown in table 3.3.

Starting from the tourist's positive or negative preferences about the region and the activities –represented as desires–, the T-agent must decide which of

Region	**Selected**	**Rejected**	**Degree(1 to 10)**
Litoral			
Atlantic coast			
Cordoba mountains			
Adean Patagonia			
Atlantic Patagonia			
Cuyo			
Northwest			

Table 3.1: T-agent: Preferred-rejected regions

Preferences	**Selected**	**Rejected**	**Degree(1 to 10)**
Sea-coast			
Mountains	X		8
Near-Places($\leq$500km)			
Middle-distant-Places			
Far away-Places($\geq$1000km)		X	9
Big cities			
Small-cities			

Table 3.2: T-agent: Requesting preferences

them the tourist may follow –modeled by intentions– through the best plan found by the Planner.

To show how our T-Agent works, we consider a particular tourist named Pedro that is looking for a one week holidays package and is willing to spend no more than 1100 pesos [3]. He hasn't got any preference about the region to go, apart from restricting its exploration range to places that are not far from Rosario, where he lives (this selection is shown in table 3.2). Pedro chooses the desires of going to a mountain place –*mountains (8)*–, visit new places –*visitNP (7)*– and in third place he chooses the activity adventure tourism, selecting *rafting(5)* (shown in tables 3.2 and 3.3, respectively). To decide an intention for Pedro, the T-Agent must look for a plan taking into account the benefit (with respect to going to the mountains, to visit new places and adventure tourism) and the cost of the proposed package. The T-Agent will consult with it's

[3] Argentine coin

Activity	Options	Degree(1 to 10)
Rest		
Visit New Places	Excursions	
Adventure Tourism		
	trekking	
	mountain-bike	
	rafting	5
Shopping		
Fiesta		

Table 3.3: T-agent: Requesting Activities

tourist supplier 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 T-agent has the following positive and negative desires:

- $(D^+(mountain), 0.8)$
- $(D^+(visitNP), 0.7)$
- $(D^+(rafting), 0.5)$
- $(D^-(distance > 1000km), 0.9)$

It is also considered the conjunction of all the single desires. According to the formalization of desires in the proposed model, the desire degree of the conjunctions are taken to be greater than their components:

- $(D^+(mountain \wedge visitNP \wedge rafting), 0.9)$
- $(D^+(mountain \wedge visitNP), 0.85)$
- $(D^+(mountain \wedge rafting), 0.82)$
- $(D^+(visitNP \wedge rafting), 0.72)$

B context:

This theory contains general knowledge about the tourist domain as the activities allowed in each place, the distances between cities, if it is a mountain or sea region, among other characteristics. Especially this theory includes the relationship between the basic actions and the plans a tourist can carry out and the formulae made true by their execution. The plans are tourist packages and may include basic actions as traveling to several destinations by different means of transport, staying in different accommodations, making various activities as excursion and sports. The basic actions and the set Π of plans used in this example: $\Pi = \{$ Atu17, Atu27, Cata17,Cata27, Cumbre17, Cumbre27, CarPaz17, CarPaz27, Bari17, ViGe17, Men17, MarPla17, PtoMa17, Ushua17, South17 $\}$, will be detailed in the Planner context.

The T-Agent has to represent also the beliefs about how the desires (e.g., mountain, visitNP, etc.) may be satisfied after executing different plans. In this example, following the model presented, it may be considered the degree of $B([\alpha]D)$ as the probability of *D* after following the plan α.

In particular the T-Agent needs to assess beliefs about the possibility that a plan, offers to visit new places–*visitNP*–. In this case the degree of $B([\alpha]visitNP)$ is interpreted as the probability of knowing new places executing the plan α. This satisfaction depends fundamentally on the destinations of the plan, and would vary slightly with the set of excursions included in it. If a tourist points out this desire, the T-Agent will ask him for the places he has been before and how long he has been in each one. This information will allow the T-Agent to evaluate the belief of *visit NP* for the different plans — as the probability that has the tourist to visit new places in each one. The satisfaction of the goal visitNP with a plan is increasing with the number of destinations that it includes.

As this is the choice made by Pedro (i.e., he has the desire visitNP), the T-agent inquiried about the places he had visited before. Considering that he had been before in Mar del Plata (20 days),Villa Gessell(15 days), Carlos Paz (15 days), Bariloche (12 days) and Mendoza(2 days); and taking into account the characteristics of each destination and the excursions included in each plan, the T-Agent estimates the corresponding beliefs. This is done using a set of rules that relate the days been in a destination with the estimation of the days needed to visit the important places in it. For this example we have the following beliefs:

- $(B([CarPazx]visitNP), 0.3)$
- $(B([MarPla1]visitNP), 0.3)$

- $(B([ViGe1]visitNP), 0.6)$
- $(B([Bari1]visitNP), 0.7)$
- $(B([Mendo1]visitNP), 0.7)$
- $(B([x]visitNP), 1)$ where x is a plan with other destination than Carlos Paz, Bariloche, Villa Gessell, Mendoza and Mar del Plata.

Considering now the desire of going to a mountain place, the degree of $B([\alpha]Mountain)$ is taked as the probability of staying in a mountain place following the plan α. Hence, the value will be 0 or 1 depending on the destination of the plan.

In this example the T-Agent has computed the following beliefs, separating the plans with destination in mountain places than other destinations:

(B([Atu17]Mountain), 1)	(B([Cata17]Mountain), 0)
(B([Atu27]Mountain),1)	(B([Cata27]Mountain), 0)
(B([Cumbre17]Mountain), 1)	(B([MarPla17]Mountain), 0)
(B([Cumbre27]Mountain),1)	(B([ViGe17]Mountain), 0)
(B([Bari17]Mountain), 1)	(B([Pto Ma17]Mountain), 0)
(B([South17]Mountain),1)	
(B([Men17]Mountain), 1)	
(B([Ushua17]Mountain),1)	

Pedro has also chosen the desire of rafting. The T-Agent has beliefs about reaching this desire after the execution of different plans, with values near to 1, if one of the destinations of the plan brings the possibility of doing this activity (the probability value will depend on the weather conditions of the destination of the plan). In order to simplify this example we consider these probabilities as 1. The value will be 0, in other cases. Considering the set of plans selected in this example, the T-Agent has the following beliefs:

- $(B([\beta]Rafting), 1)$ if β is a plan with one destination in {Atuel Canyon, Bariloche, Pto. Iguazu or Mendoza }.
- $(B([\alpha]Rafting), 0)$ otherwise.

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

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

P Context

In this context the T-Agent finds the tourist plans starting from a set of elementary actions (travel to a destination, staying in several accommodation, making excursions, sports, etc.), formally:
$action(description, P, A)$ where P are the pre-conditions and A state for the post-condition of the actions.

The T-Agent has the following set of basic actions:

- $action(travel(destination), [cost], [stay(destination, accommodation, days])$ where $travel \in \{travelbybus, travelbytrain, travelbyowncar, fly\}$; and destination is one of the listed places, for example:
- $action(stay(destination, accommodation, xdays), [travel(destination), cost], [])$ where $accommodation \in \{camping, hotel3*, hotel4*, hotel5*, bungalow, hostel\}$
- $action(excursion(excursion_i), [cost, stay(destination, accommodation, xday)], [tired, visitNP])$ where $excursion_i$ is able from destination.
- $action(sport(sport_j), [cost, stay(destination, accommodation, xday)], [tired, healthy])$ where $sport_j$ is possible in destination.

Example of concrete actions are:

- *action(travelbyplane(Calafate),[cost(800)], [stay(Calafate, accommodation, days])*
- *action(stay(Calafate, hotel3*,3days), [travel(Calafate),cost(240)],[ ])*
- *action (sport(glaciertrekking),[cost(100),stay(Calafate, accommodation, xday)],[tired, visitNP])*

The T-Agent looks for suitable plans. Each one may include one or more destinations, accommodations, activities as excursions and sports, depending on the interest of the tourist, and the packages offered by the tourism provider.

A plan then, has a name or label, as preconditions has it cost, and the postconditions are a concatenation of different actions the plan includes. We added for the intention calculus the term Cn, the normalized cost: $Cn = cost/MaxCost$, where MaxCost is the maximum cost the tourist is willing to spend. The Planner will also use, for each goal trying to satisfy, the belief degree b of the goal satisfaction through the plan, namely $b = B[plan]goal$, this will be use to filter the plans which belief degree b is greater or equal than some threshold, called $b - threshold$. Then plans are represented by the following tuple:

- plan(label, [cost], [travel(destination1), stay(destination1, accommodation1, x1days),excursion(excursion1i), sport(sport1j),...,travel(destinationj), stay(destinationj, accommodationj, xjdays),excursion(excursionji), sport(sportjk)], Cn)

For instance:

- plan(Atu17, [cost(600)],[bus(AtuelCanyon), stay(AtuelCanyon, bungalow,7days), sport(rafting)], Cn)
- plan(South17,[cost(1500)],[fly(Calafate), stay(Calafate, hotel 3*, 3days), excursion(PeritoMorenoGlacier), sport(glaciertrekking), fly(Ushuaia), stay(Ushuaia, hotel 3*, 4days), excursion (TierradelFuegoChannels)], Cn)

Once these theories are defined, the T-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 of resting, visiting new places and rafting, through the different actions. Using the restriction introduced by the negative desire: $(D^{-}(dist > 1000km), 0.9)$ the planner rejects plans to Patagonia region (Bariloche, Pto Madryn, Calafate, Usuhaia) and to Northwest region because their post-conditions of stay in these destinations make true $(dist > 1000km)$ which is strongly rejected (0.9). The b-threshold is set in 0.5 and is used to filter all the plans α such that $B([\alpha]\varphi) < 0.5$, for each desire φ, and then do not satisfy $(B([\alpha]\varphi), b - threshold)$. Considering that Pedro is willing to spend no more than 1100 pesos, the agent sets MaxCost=1100 pesos for this example. Therefore, using the bridge rule

$$\frac{D : \nabla(D^{+}\varphi), D : (D^{-}\psi, \mathit{threshold}), P : \mathit{action}(\alpha, P, A, c), \quad B : (B([\alpha]\varphi), \mathit{bthreshold}), B : B(A \rightarrow \neg\psi)}{P : \mathit{plan}(\varphi, \alpha, P, A, c)}$$

plans are generated for each desire D: $plan_D$. In what follows we show some of them:

a- For instance, for the most preferred positive desire, i.e. $D_1 = mountain \wedge visitNP \wedge rafting$, the following plans are generated:

- $plan_{D_1}$(Atu17,[cost(600)],[bus(AtuelCanyon), stay(AtuelCanyon, bungalow,7days), sport(rafting)], Cn=0.545),
  where $Cn = cost/MaxCost = 600/1100$
- $plan_{D_1}$(Atu27,[cost=800],[bus(AtuelCanyon), stay(AtuelCanyon, hotel3*,7days), excursion(Canyon), sport(rafting)],Cn=0.727)
- $plan_{D_1}$(Mendo17,[cost=600],[bus(Mendoza), stay(Mendoza, hotel3*,7days), excursion(LasLeas)],Cn=0.545)

b- Considering $D_2 = mountain \wedge visitNP$ the following plans are generated:

- $plan_{D_2}$(Atu17,[cost=600],[bus(AtuelCanyon),stay(AtuelCanyon, bungalow,7days), sport(rafting)], Cn=0.545)
- $plan_{D_2}$(Atu27,[cost=800],[bus(AtuelCanyon), stay(AtuelCanyon, hotel3*,7days), excursion(Canyon), sport(rafting)], Cn=0.727)
- $plan_{D_2}$(Mendo17, [cost=600],[Mendo17,[cost=700],[bus(Mendoza), stay(Mendoza, hotel3*,7days), excursion(LasLeas)],Cn=0.636)
- $plan_{D_2}$(Cumbre17, [cost=400],[bus(Cumbrecita),stay(Cumbrecita, hotel2*,7days)], Cn=0.364)
- $plan_{D_2}$(Cumbre27, [cost=500],[bus(Cumbrecita),stay(Cumbrecita, hotel3*,7days)], Cn=0.455)

Notice that the plans CarPaz17 and CarPaz27 are filter by the bridge rule, because $b = B([CarPazj]D_2) = B([CarPazj]mountain) \cdot (B[CarPazj]visitNP) < 0.5$.

c- For $D_3 = mountain \wedge rafting$ and $D_4 = visitNP \wedge rafting$ the same plans than for D_1 are generated, because of the restriction imposed by the desire of rafting (varying in the belief of the satisfaction of the different goals by the diverse plans).

3. *The plans determine the degree of intentions*

The intention degree depends on the benefit and the cost of reaching a goal. Using bridge rule

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

and the function $f(d, c) = (d + (1 - c))/2$ proposed for an *equilibrated* agent the I context calculates the intention degree for the different destinations.

a- Hence, $D_1 = mountain \wedge visitNP \wedge rafting$ is preferred to a degree 0.9, using $f(d, c) = (0.9 + (1 - c))/2$ we successively have for the different plans generated for this composed desire D_1 ($I_{[\alpha]}$ denotes the intention through the plan α):

$(I_{[Atu17]}D_1, 0.678)$,
$(I_{[Atu27]}D_1, 0.587)$,
$(I_{[Mendo17]}D_1, 0.632)$,

We get a maximal degree of intention for $I(mountain \wedge visitNP \wedge rafting)$ of 0.678, corresponding to the plan $Atu17$.

b- Considering $D_2 = mountain \wedge visitNP$, desire preferred in degree 0.85, and using the function $f(d, b) = (0.85 + (1 - c))/2$ the T-Agent has successively the following intentions degrees, for the different plans $\alpha \in \Pi_{D2}$:

$(I_{[Atu17]}D_2, 0.653)$,
$(I_{[Atu27]}D_2, 0.562)$,

$(I_{[Cumbre17]}D_2, 0.743)$,
$(I_{[Cumbre27]}D_2, 0.698)$,
$(I_{[Mendo17]}D_2, 0.607)$,

We get a maximal degree of intention for $D_2 = mountain \wedge visitNP$ of 0.743 for the plan Cumbre17 ($I_{[Cumbre17]}mountain \wedge visitNP = 0.743$).

c- Considering $D_3 = rest \wedge rafting$ (with degree 0.82) and $D_4 = visitNP \wedge rafting$ (degree 0.71) T-Agent have selected the same plans than for D_1. Since f is monotonically increasing with respect to d, for the same set of plans, it is enough had considered the most preferred desired, i.e. $mountain \wedge visitNP \wedge rafting$.

4. *A plan is adopted*

   Finally, by means of bridge rule

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

   the plan $\alpha = Cumbre17$ is selected and passed to the Communication context CC.

   Notice that in this case, the T-Agent selected the plan which brings the higher intention degree corresponding to reach a goal D_2 which is less desired than D_1. Depending on the application's requirement this selection may be changed, choosing a suitable function f for each case.

3.8 Conclusions

In this exploratory work we have presented a BDI agent model that allows to explicitly represent the uncertainty of beliefs, degrees in desires and intentions. This graded architecture is specified using multi-context 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 Łukasiewicz multivalued logic is the framework chosen to formalize the degrees and we added the corresponding axiomatic in order to represent the uncertainty behavior as probability, necessity and possibility. Other measures of uncertainty might be used in the different units by simply changing the corresponding axiomatic. Also, the model introduced, based on a multi-context specification, can be easily extended to include other mental attitudes. Adding concrete theories to each context, particular agents may be defined using our context framework. The agent's behavior is then determined by the different uncertainty measures of each context, the specific theories established for each unit, and the bridge rules.

For future work, we are considering two main directions. On the one hand we want to extend our multi-context agent model to a multiagent scenario. On the other hand, from an computational point of view, our idea is to implement this BDI graded model. 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. The future work is outlined in the Thesis Proposal of the following Chapter.

Chapter 4

Thesis Project

4.1 Proposal

The doctoral thesis project is in the field of agent's architectures and in particular in those based in the intentional stance, as the BDI model. We have begun in this exploratory work with the individual aspect of agency, proposing a graded BDI model for agents. This model allows to represent and reason using degrees in the belief, desire and intention's attitudes. We consider that the new model has a richer semantics with respect to the classic BDI approaches (e.g., the BDI architecture by Rao and Georgeff described in Section 2.1.3 or the multi-context BDI agent showed in Section 2.1.4). Moreover, the multi-context specification used in our model, offers a clear path from specification to implementation (this was discussed at the end of Section 2.1.4). Based in this first approach, we consider that graded models for agent's architectures will be a contribution to the develop of more flexible agents capable of taking better decisions. To achieve this goal, there are some challenges for the next future, we focus the thesis in some of them.

The model must be extended in order to incorporate other important aspects related to the BDI model as belief revision and intention reconsideration, among others. Furthermore, the social aspects of agency must be faced up with the purpose of embedding the BDI graded model in a multiagent platform. Besides, a prototype implementation is planned. This will allow us on the one hand, to specified particular agents with different behaviors, and on the other hand, to experiment with the model.

In what follows, a working plan is proposed to achieve the general objective of this thesis: to make its contribution to the development of new models for agent's architectures, including uncertain beliefs and graded pro-attitudes. Particularly we want to specify a *social-graded BDI agent* capable to interact intelligently in a multiagent platform.

4.2 Working Plan

In this Section we outline the main stages of the future research work. This work will take place mainly in the Departamento de Sistemas e Informática (DSeI), Facultad de Cs. Exactas, Ingeniería y Agrimensura (FCEIA), Universidad Nacional de Rosario (UNR), Argentine, where I work as a professor. We plan for each year, a one month stay at the Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC), Spain, to work in close collaboration with the research work directors. The principal tasks to be developed are the following:

1. **Refine and extend the proposed model for graded agent's architectures.**
   This must be done in different directions, we following mention some of them. We can experiment different agent's behaviors choosing other measures of uncertainty to be used in the Belief, Desire and Intention units, by changing the corresponding axiomatic or the chosen logic. Also, the model introduced, based on a multi-context specification, can be extended to include other mental attitudes as obligations, commitments, etc. Another important points to consider in our model are, the implementation of a belief revision process, to maintain the consistence of the agent's beliefs; and an intention reconsideration policy, to establish a good balance between action and deliberation. Some of this process will be analyzed during a concrete application, as are strongly related to the domain's characteristics.

2. **Develop the social aspects of the graded BDI agent model.**
   An extension of our multi-context agent model to a multiagent scenario is needed. 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. We consider important to filter the information interchange among agents, as it is properly approached in [1]. 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 to

be consider in our work are Liau's logic of Belief, Information and Trust (BIT) [47] and the extension of this model described in [19].

3. **Define specific agent's architectures using this model.**
   Adding concrete theories to each context, particular agents may be defined using our multi-context blueprint. The agent's behavior is then determined by the different uncertainty measures of each context, the specific theories established for each unit, and the bridge rules. We are planning to explore the tourist domain, in order to obtain a chain of tourist-providers agents (the travel-agency; the tourist packages provider in a middle level, and the air companies, hotels, etc., in the higher level). Different works about Recommender Agents will be taken into account for this application and in particular, the approach made by B. Lopez in tourism [49] will be considered. The tourist domain seems to be a rich application context where we can experiment either with different individuals agent's behaviors, or with socials relations.

4. **Implementation.**
   An implementation using Prolog language is planned. This will follow from an implementation of multi-context Systems developed by Andrea Giovannucci [29]. Our idea is to implement each unit as a 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.

5. **Experimentation and evaluation.**
   We plan the first prototype in the tourist domain but we also plan other possible implementation in Health Care. In particular, in the *Sistemas Inteligentes* group (FCEIA-UNR) we are working together with the *Cátedra de Endodoncia* (Facultad de Odontología-UNR) in the development of intelligent systems to support the diagnosis education [11]. In addition, between our group and the Agent Research Laboratory of the Universidad de Girona (UdG), began a Cooperation Project to develop "Knowledge based decision support systems for diagnosis and coordination of health services"

[1]. Also, the Artificial Intelligence Research Intitute (IIIA - CSIC) has long experience in applications of medical expert systems [33] as well as, more recently, in multi-agent systems in the Health Care environment [34]. We think that these cooperations in the Health Care field will bring us new application domains, were a graded agent approach will be useful; because health domain is naturally uncertain, complex and dynamic. The different implementations will allow us to experiment and validate the formal model presented.

6. **Documentation.**
We will report the partial results and main contributions in international forums, such as conferences and journals related to this research and application area. And finally, the PhD dissertation will be prepared and documented.

We schematize a working plan of thirty months, divided in three-months period (Ti). The proposed schedule for the main stages previously enumerated, is summarized in the following flow-chart:

Activity description	Time Line									
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
1. Refine and extend the proposed model										
2. Develop the social aspects of the model										
3. Define specific agent architectures										
4. Implementation: general model and specific agents										
5. Experimentation and evaluation										
6. Documentation										

Table 4.1: Schedule for the doctoral research work (Ti=Three-months period)

[1]Programa de Cooperación Interuniversitario con Iberoamérica (Intercampus), supported by AECI-MAE, Spain, january 2005 - january 2006

Bibliography

[1] Acebo, E. And de la Rosa, J.LL. A Fuzzy System Based Approach to Social Modeling in multiagent Systems. *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems* (AAMAS'02), Palazzo Re Enzo, Italy, 2002.

[2] M. Benerecetti, F. Giunchiglia, and L. Serafini. Model Checking Multiagent Systems. *Journal of Logic and Computation, Special Issue on Computational and Logical Aspects of Multi-Agent Systems*, 8(3):401-423, 1998.

[3] Benferhat S., Dubois D. and Prade, H. Towards a possibilistic logic handling of preferences. *Applied Intelligence* 14(3):303317, 2001.

[4] 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.

[5] 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.

[6] C. Boutilier. Towards a logic for qualitative decision theory. In *Proceedings of International Conference on Principle of Knowledge Representation and Reasoning* (KR94), 7586, 1994.

[7] Bratman, M. E. *Intention, Plans and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[8] M. E. Bratman, D. J. Israel, and M. E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349355, 1988.

[9] Bratman, M. E. What is intention? *In Intention in Communication* (editors P.R. Cohen, J.L. Morgan and M.E. Pollak) pp. 15-32. MIT University Press, Cambridge, MA, 1990.

[10] Brooks, R. A. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2(1):14-23, 1986.

[11] Casali, A., Corti R., DAgostino E. y Siragusa M. Herramienta Tecnolgica como apoyo al Diagnstico Endodntico. *Electronic Journal of Endodontics*, EJER (ISSN 1666-6143) , year 3, vol 1: 10 pages. Rosario, Argentine, 2004.

[12] Casali A., Godo Ll. and Sierra C. Modelos BDI para arquitecturas Graduadas de Agentes. In *Proceedings of the 6th Argentine Symposium on Artificial Intelligence* (ASAI 2004- 33JAIIO) (ISSN 1666 1079) 13 pages, Crdoba, Argentina, 2004.

[13] Casali A., Godo Ll. and Sierra C. Modelos BDI para arquitecturas Graduadas de Agentes. Next to appear in an special number of *Artificial Intelligence, Iberoamerican Publication, AEPIA* (ISSN: 1137-3601 (c))

[14] Casali A., Godo Ll. and Sierra C. Graded Models for BDI Agents. In Proceedings of the *Fifth International Workshop on Computational Logic in Multi-agent Systems* (CLIMA V) , Leite J. and Torroni P. eds., 18-33, ISBN:972-9119-37-6. Lisboa, Portugal, 2004. Next to apear in *Post-Proceedings of CLIMA V, in an special Volumen of the Lectures Notes in Artificial Intelligence* (LNAI), Springer-Verlag.

[15] Casali A., Godo Ll. and Sierra C. Graded Models for BDI Agents. In *Proceedings of European Workshop of Multiagent Systems* (EUNAS'04), Barcelona, Spain, 2004.

[16] 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, number 890 in Lecture Notes in Computer Science*, pages 71-5. Springer Verlag, 1995.

[17] Cohen, P. R. and Levesque, H. J. Intention is choice with commitment. *Artificial Intelligence*, 42:213-261, 1990.

[18] Cohen, P. R. and Levesque, H. J. Communicative actions for artificial agents. In Proceedings of the *First International Conference on Multi-Agent Systems* (ICMAS-95), pages 65–72, Menlo Park, CA, AAAI Press, 1995.

[19] 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. eds),34-49, 2004.

[20] Dennet, D. C. *The Intentional Stance*. MIT Press, Cambridge, MA, 1987.

[21] d'Inverno M. et al. A Formal Specification of dMARS. In *Intelligent Agents, IV* (eds Rao A., Singh P. and Wooldridge M.), LNAI Volume 1365, pp 155-176, Springer, Berlin, 1997.

[22] Dubois D. and Prade H. Possibility Theory: qualitative and quantitative aspects. In *Hanbook of Defeasible Reasoning and Uncertainty Management Systems* (eds. Gabbay D. and Smet Ph.), Vol I: Quantified Representations of Uncertainty and Imprecision, 169-226,Kluwer, 1998.

[23] Emerson E. A. and Strinivasan J. Branching time temporal logic. In J. W. de Bakker, W. P. de Roever and G. Rozenberg, editors. *Linear Time, Branching Time and Partial Order in Logics and Models for Concurrency*, pages 123-172. Springer -Verlag, 1989.

[24] Ferguson, I. A. TouringMachines: an Architectures for Dynamic, Rational , Mobile Agents. *PhD thesis, Clare Hall, University of Cambrage, UK.* (Also available as Technical Report n 273, University of Cambrage Computer Laboratory).

[25] M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87), pages 677682, Seattle, WA, 1987.

[26] M. P. Georgeff and A. S. Rao. A profile of the Australian AI Institute. IEEE Expert, 11(6):8992, December 1996.

[27] M. Georgeff, B. Pell, M. Pollack, M. Tambe, and M. Wooldridge. The Belief-Desire-Intention Model of Agency. In *Intelligent Agents V Springer-Verlag Lecture Notes in AI* (J. P. Muller, M. Singh, and A. Rao, eds.) Volume 1365, March 1999.

[28] Ghidini C. and Giunchiglia F. Local Model Semantics, or Contextual Reasoning = Locality + Compatibility *Artificial Intelligence*,127(2):221-259, 2001.

[29] Giovannucci A. Towards Multi-Context based Agents Implementation. *IIIA-CSIC Research Report*, 2004.

[30] Giunchiglia F. Contextual Reasoning. In *Proceedings of the IJCAI* Workshop on Using Knowledge in Context, 1993.

[31] 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.

[32] Godo, L., Esteva, F. and Hajek, P. Reasoning about probabilities using fuzzy logic. *Neural Network World*, 10:811–824, 2000.

[33] Godo L., López de Mántaras R., Puyol-Gruart J., Sierra C. Renoir, Pneumon-IA and Terap-IA: three medical applictions based on fuzzy logic. *Artificial Intelligence in Medicine* 21 (2001) 153-162.

[34] Godo L., Puyol-Gruart J., Sabater J., Torra V., Barrufet P., Fbregas X. A multi-agent system approach for monitoring the prescription of restricted use antibiotics. *Artificial Intelligence in Medicine* 27 (2003) 259282.

[35] Goldblatt R. *Logics of Time and Computation*, CSLI Lecture Notes 7, 1992.

[36] Hájek P., Godo L. and Esteva F. Fuzzy Logic and Probability. In *Proceedings of the 11th. Conference Uncertainty in Artificial Intelligence 95* (UAI'95) ,pp. 237-244, 1995.

[37] Hájek, P. *Metamathematics of Fuzzy Logic*, volumen 4 de Trends in Logic. Kluwer, 1998.

[38] Huber, M. Jam: a BDI-theoretic mobile agent architecture. In *Proceedings of the 3rd International Conference on Autonomous Agents* (Agents 99), Seatle, WA, 236-243, 1999.

[39] http://www.agentlink.org

[40] http://www.agentland.com

[41] http://www.insead.fr/CALT/Encyclopedia/ComputerSciences/Agents/

[42] http://www.agentcities.net

[43] Ingrand F. F., Georgeff M. P. and Rao A. S. An architecture for real time reasoning and system control. *IEEE Expert*, 7(6), 1992.

[44] Jennings N.R. On Agent-Based Software Engineering. *Artificial Intelligence* 117(2), 277-296, 2000.

[45] Lang J. Conditional Desires and Utilities - an alternative logical approach to qualitative decision theory. In *Proceedings of 12th European Conference on Artificial Intelligence* (ECAI'96), 318-322, 1996.

[46] 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.

[47] Liau C. J., Belief, Information Acquisition, and Trust in Multiagent Systems - a modal formulation. *Artificial Intelligence* 149, 31-60, 2003.

[48] Ljungberg M. and Lucas A. The OASIS Air-traffic Management System. *In Proceedings os the Second Pacific Rim International Conference on Artificial Intelligence* (PRICAI92), Scoul, Korea, 1992.

[49] López B. Holiday Scheduling for City Visitors. *Information and Communication Technologies in Tourism*, A.J. Frew, M. Hitz, P. OConnor, (eds.),pginas 252-260, Springer Verlag, 2003 ENTER03, Helsinki.

[50] López de Mántaras R. *Approximate Reasoning Models*. Ellis Horwood, 1990.

[51] Maes P. The Agent netword architecture (ANA). *SIGART Bulletin*, 2(4) 115-120, 1991.

[52] Luck M., McBurney P., and Preist C. Agent Technology: Enabling Next Generation Computing. *AgentLink, 2003*. ISBN 0854 327886. (http://www.agentlink.org/roadmap/)

[53] Meyer J. J. Dynamic Logic for Reasoning about Actions and Agents. *Workshop on Logic-Based Artificial Intelligence*, Washington, DC, June 14-16, 1999.

[54] Muller J.P., Pischel, M. and Thiel M. Modeling Reactive Behaviour in Vertically Layered Agent Architectures. In *Intelligent Agents: theories, Architectures and Languages* (eds. Wooldridge M. and Jennings N.), LNAI Valume 890, pp. 261-276. Springer, Berlin, 1995.

[55] Muller J.P. A cooperation Model for Autonomous Agent. In *Intelligent Agents, III*(eds. J. P. Muller, M. Wooldridgde and N. R. Jennings), LNAI Volume 1193, pp245-260, Springer, Berlin, 1997.

[56] Noriega P. and Sierra C.Towards layered dialogical agents. In *Intelligent Agents III* (Muller J. P., Wooldridge M. J., and Jennings N. R., eds), pp.173-188, Springer Verlag, Berlin, 1996.

[57] Parsons, S., Sierra, C. and Jennings N.R. Agents that reason and negociate by arguing. *Journal of Logic and Computation*, 8(3): 261-292, 1998.

[58] 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.

[59] Parsons S., Pettersson O., Saffiotti A. and Wooldridge M. Intention Reconsideration in Theory and Practice. In Proceedings of the *14th European Conference on Artificial Intelligence*, (ECAI 2000) IOS Press : 378-382,Berlin, Germany, 2000.

[60] Parsons,S., Jennings,N.J., Sabater,J. and Sierra C. Agent Specification Using Multi-context Systems. *Foundations and Applications of Multi-Agent Systems* 205-226, 2002.

[61] 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.

[62] Rao, A. And Georgeff M. Deliberation and Intentions. In *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Mateo, CA, 1991.

[63] 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.

[64] A. S. Rao and M. Georgeff. Decision procedures of BDI logics. *Journal of Logic and Computation*, 8(3):293344, 1998.

[65] Russell, S. and Norvig P. *Artificial Intelligence: a modern approach*. Prentice Hall. Englewoods Cliifs, NJ, 1995.

[66] 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.

[67] Sabater, J. PhD Thesis: Trust and Reputation for Agent Socities. Artificial Intelligence Institute Monograph, IIIA-CSIC N20, 2003.

[68] K. Schild. On the relationship between BDI logics and standard logics of concurrency. In J. P. Muller, M. P. Singh, and A. S. Rao, eds, Intelligent Agents V Proceedings of the *Fifth International Workshop on Agent Theories, Architectures, and Languages (ATAL-98)*, Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg, 1999.

[69] 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.

[70] Schut, M., Wooldridge, M. and Parsons S. Reasoning About Intentions in Uncertain Domains. In Proceedings of *Symbolic and Quantitative Approaches to Reasoning with Uncertainty. 6th European Conference* (ECSQARU 2001), 84-95, Toulouse, France, 2001.

[71] van der Torre L. and Weydert E. Parameters for Utilitarian Desires in a Qualitative Decision Theory. *Applied Intelligence*, 14:285-301, 2001.

[72] Weiss, G.(editor) *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*,The MIT Press, 1999.

[73] Wooldridge, M. and Jennings N.R. Intelligent Agents: theory and practice. *The Knowledge Engineering Review*, 10(2), 115-152, 1995.

[74] Wooldridge, M. and Jennings N.R. Agent-based software engineering. *IEEE Proceedings in Software Engineering*, 144(1), 26-37, 1997.

[75] Wooldridge, M. and Parsons S.: Intention Reconsideration Reconsidered. In Proceedings of *Intelligent Agents V, Agent Theories, Architectures, and Languages, 5th International Workshop* (ATAL '98), Paris, France, 1998. *Lecture Notes in Computer Science 1555 Springer* (Jrg P. Mller, Munindar P. Singh, Anand S. Rao Eds.) ISBN 3-540-65713-4A, 63-79, 1999.

[76] Wooldridge, M. *Introduction to Multiagent Systems*, John Wiley and Sons,Ltd., 2001