An approach to using degrees of belief in BDI agents

Simon Parsons and **Paolo Giorgini**[*]
Department of Electronic Engineering,
Queen Mary and Westfield College,
University of London, London E1 4NS, UK.

Abstract: The past few years have seen a rise in the popularity of the use of mentalistic attitudes such as beliefs, desires and intentions to describe intelligent agents. Many of the models which formalise such attitudes do not admit degrees of belief, desire and intention. We see this as an understandable simplification, but as a simplification which means that the resulting systems cannot take account of much of the useful information which helps to guide human reasoning about the world. This paper starts to develop a more sophisticated system based upon an existing formal model of these mental attributes.

1 Introduction

In the past few years there has been a lot of attention given to building formal models of autonomous software agents; pieces of software which operate to some extent independently of human intervention and which therefore may be considered to have their own goals and the ability to determine how to achieve those goals. Many of these formal models are based on the use of mentalistic attitudes such as beliefs, desires and intentions. The beliefs of an agent model what it knows about the world, the desires of an agent model those states of the world the agent finds preferable, and the intentions of an agent model those states of the world that the agent actively tries to bring about. One of the most popular and well-established of these approaches is the BDI model of Rao and Georgeff [17, 18].

While Rao and Georgeff's model explicitly acknowledges that an agent's model of the world is incomplete, by modelling beliefs as a set of worlds which the agent knows that it might be in, the model makes no attempt to make use of information about how likely a particular possible world is to be the actual world in which the agent operates. Our work is aimed at addressing this issue, which we feel is a weakness of the BDI model, by allowing an agent's beliefs, desires, and intentions to be quantified. In particular this paper considers quantifying an agent's beliefs using Dempster-Shafer theory, which immediately makes it possible for an agent to express its opinion on the reliability of the agents it interacts with and to revise its beliefs when they become

[*]Current address: Dipartimento di Informatica e Studi Aziendali, Università degli Studi di Trento, Via Inama 5, I-38100 Trento, Italy.

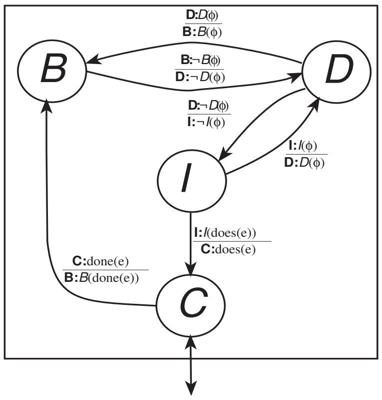

Figure 1: The multi-context representation of a strong realist BDI agent

inconsistent. To do this, the paper combines the first author's work on the use of argumentation in BDI agents [14], with the second author's work on belief revision [4]. The question of quantifying desires and intentions is the subject of continuing work.

2 Multi-context agents

As mentioned above, our work here is an extension of that in [14] to include degrees of belief. As in [14] we describe our agents using the framework of multi-context systems [10]. We do this not because we are interested in explicitly modelling context, but because multi-context systems give a neat modular way of defining agents which is then directly executable. This section briefly recaps the notions of multi-context systems and argumentation as used in [14].

Using the multi-context approach, an agent architecture consists of the following four components (see [13] for a formal definition):

- *Units*: Structural entities representing the main components of the architecture. These are also called *contexts*.
- *Logics*: Declarative languages, each with a set of axioms and a number of rules of inference. Each unit has a single logic associated with it.
- *Theories*: Sets of formulae written in the logic associated with a unit.
- *Bridge rules*: Rules of inference which relate formulae in different units.

The way we use these components to model BDI agents is to have separate units for belief B, desires D and intentions I, each with their own logic. The theories in each unit encode the beliefs, desires and intentions of specific agents, and the bridge rules encode the relationships between beliefs, desires and intentions. We also have a unit C which handles communication with other agents. Figure 1 gives a diagrammatic representation of this arrangement. For each of these four units we need to say what the logic used by each unit is. The communication unit uses classical first order logic with the usual axioms and rules of inference. The belief unit also uses first order logic, but with a special predicate B which is used to denote the beliefs of the agent. Under the modal logic interpretation of belief, the belief modality is taken to satisfy the axioms K, D, 4 and 5 [19]. Therefore, to make the belief predicate capture the behaviour of this modality, we need to add the following axioms to the belief unit (adapted from [2]):

$$\begin{array}{ll}
\mathbf{K} & B : B(\varphi \to \psi) \to (B(\varphi) \to B(\psi)) \\
\mathbf{D} & B : B(\varphi) \to \neg B(\neg\varphi) \\
\mathbf{4} & B : B(\varphi) \to B(B(\varphi)) \\
\mathbf{5} & B : \neg B(\varphi) \to B(\neg B(\varphi))
\end{array}$$

The desire and intention units are also based on first order logic, but have the special predicates D and I respectively. The usual treatment of desire and intention modalities is to make these satisfy the K and D axioms [19], and we capture this by adding the relevant axioms. For the desire unit:

$$\begin{array}{ll}
\mathbf{K} & D : D(\varphi \to \psi) \to (D(\varphi) \to D(\psi)) \\
\mathbf{D} & D : D(\varphi) \to \neg D(\neg\varphi)
\end{array}$$

and for the intention unit:

$$\begin{array}{ll}
\mathbf{K} & I : I(\varphi \to \psi) \to (I(\varphi) \to I(\psi)) \\
\mathbf{D} & I : I(\varphi) \to \neg I(\neg\varphi)
\end{array}$$

Each unit also contains the *generalisation*, *particularisation*, and *modus ponens* rules of inference. This completes the specification of the logics used by each unit.

The bridge rules are shown as arcs connecting the units. In our approach, bridge rules are used to enforce relations between the various components of the agent architecture. For example the bridge rule between the intention unit and the desire unit is:

$$I : I(\alpha) \Rightarrow D : D(\lceil \alpha \rceil) \tag{1}$$

meaning that if the agent has an intention α then it desires α[1]. The full set of bridge rules in the diagram are those for the "strong realist" BDI agent discussed in [19]:

$$\begin{array}{rcll}
D : \neg D(\alpha) & \Rightarrow & I : \neg I(\lceil \alpha \rceil) & (2) \\
D : D(\alpha) & \Rightarrow & B : B(\lceil \alpha \rceil) & (3) \\
B : \neg B(\alpha) & \Rightarrow & D : \neg D(\lceil \alpha \rceil) & (4) \\
C : done(e) & \Rightarrow & B : B(\lceil done(e) \rceil) & (5) \\
I : I(\lceil does(e) \rceil) & \Rightarrow & C : does(e) & (6)
\end{array}$$

Note that in the remainder of the paper we drop the 'B :', 'D :' and 'I :' to simplify the notation.

The meaning of most of these rules is obvious. The two which require some additional explanation are (5) and (6). The first is intended to capture the idea that

[1] When a term from one context comes into the scope of another, for instance because of the action of a bridge rule, it needs to be quoted using $\lceil \cdot \rceil$.

if the communication unit obtains information that some action has been completed (signified by the term $done$) then the agent adds it to its set of beliefs. The second is intended to express the fact that if the agent has some intention to do something (signified by the term $does$) then this is passed to the communication unit (and via it to other agents). It should be noted that this formalization ignores the temporal aspects in [19]—all the α in the above refer to future events, so that, for instance, if the agent does not believe α can become true in the future, it is prevented from desiring that it becomes true.

With these bridge rules, the shell of a strong realist BDI agent is defined in our multi-context framework. To complete the specification of a complete agent it is necessary to fill out the theories of the various units with domain specific information, and it may be necessary to add domain specific bridge rules between units. For an example, see [14].

3 Multi-context argumentation

The system of argumentation which we use here is based upon that proposed by Fox and colleagues [7, 11]. As with many systems of argumentation, it works by constructing a series of logical steps (arguments) for and against propositions of interest and as such may be seen as an extension of classical logic. In classical logic, an argument is a sequence of inferences leading to a true conclusion. In the system of argumentation adopted here, arguments not only prove that propositions are true or false, but also suggest that propositions might be true or false. The strength of such a suggestion is ascertained by examining the propositions used in the relevant arguments. This form of argumentation may be seen as a formalisation of work on informal logic and argumentation in philosophy [22], though it should be stressed that it was developed independently.

We fit argumentation into multi-context agents by building arguments using the rules of inference of the various units and the bridge rules between units. The use we make of argumentation is summarised by the following schema:

$$\Gamma \vdash (\varphi : G : \alpha)$$

where:

- Γ is the set of formulae available for building arguments;
- $\vdash$ is a suitable consequence relation;
- φ is the proposition for which the argument is made;
- G, the grounds, indicates the set of formulae used to infer φ; and
- α is the degree of belief (also called the "credibility") associated with φ as a result of the deduction.

This kind of reasoning is similar to that provided by labelled deductive systems [9], but it differs in its use of the labels. Whilst most labelled deductive systems use their labels to control inference, this system of argumentation uses the labels to determine which of its conclusions are most valid. For a lengthier discussion of the relationship between labelled deductive systems and our model of argumentation, see [8].

With this in mind, we can formally define an argument in our framework:

Definition 1 *Given an agent a, an* argument *for a formula φ in the language of a is a triple $(\varphi : P : \alpha)$ where P is a set of grounds for φ and α is the degree of belief in φ suggested by the argument.*

It is the grounds of the argument which relate the formulae being deduced to the set of formulae it is deduced from:

Definition 2 *A set of* grounds *for φ in an agent a is an ordered set $\langle s_1, \ldots, s_n\rangle$ such that:*

1. *$s_n = \Gamma_n \vdash_{d_n} \varphi$;*
2. *every s_i, $i < n$, is either a formula in the theories of a, or $s_i = \Gamma_i \vdash_{d_i} \psi_i$;*
3. *$d_i = a_{\{r_1,\ldots,r_m\}}$ means that the formula φ is deduced by agent a from the set of formulae Γ by using the set of inference rules or bridge rules $\{r_1, \ldots, r_m\}$ (when there is no ambiguity the name of the agent will be omitted); and*
4. *every p_j in every Γ_i is either a formula in the theories of agent a or ψ_k, $k < i$.*

We call every s_i a step *in the argument.*

For the sake of readability, we often refer to the conclusion of a deductive step with the identifier given to the step. As an example of how arguments are built, consider the following.

Example 1 If we have an agent k which is equipped with propositional logic and the theory $\{a \wedge b\}$ then it would have an argument $(a : \langle\{a \wedge b\} \vdash_{k_{\{\wedge-elimination\}}} a\rangle : \alpha)$ for some degree of credibility α. If, instead, k had the theory $\{a, a \rightarrow b, b \rightarrow c\}$, then it would have an argument $(c : \langle s_1, s_2\rangle : \beta)$ where $s_1 = \{a, a \rightarrow b\} \vdash_{k_{\{mp\}}} b$, and $s_2 = \{b, b \rightarrow c\} \vdash_{k_{\{mp\}}} c$. $\square$

At first sight, it might seem that our formulation of an argument is rather bulky. It certainly involves more baggage than other systems where, for instance, there is no explict record of the steps in the proof or the rules of inference rules used—as, for example, in the systems discussed by Dung [6], Loui [12] and Pollock [15]. The reason for this difference is that other authors have been concerned with argumentation in a single agent, and that agent can therefore reconstruct the argument for any proposition from a simple list of the propositions used. In our case, because a number of agents are involved, there is no guarantee that any one argument can be reconstructed by a given agent, and so we explicitly record every detail about the way in which the argument was built.

4 Adding degrees of belief

In our previous work we have considered agents whose belief, desire and intention units contain formulae of the form:

$$B(\varphi) \wedge B(\varphi \rightarrow \psi) \rightarrow B(\psi)$$

These have then been used to build arguments as outlined in the previous section. What we want to do is to permit the beliefs, desires and intentions to admit degrees, so that beliefs can have varying degrees of credibility, desires can be ordered, and intentions adopted with varying degrees of resolution. Since argumentation already allows us to incorporate degrees of belief it is reasonably straightforward to build in this component, and doing so is the subject of the rest of this paper. Handling degrees of desire and intention is more problematic, and this is the subject of continuing work, in particular work which associates utilities with desires and expected utilities with intentions (these expected utilities are based upon the utility of the state which the intention aims to bring about and the probability that adopting the intention will achieve that state). This work aims to build upon the earlier work of Rao and Georgeff in relating beliefs, desires and intentions to decision theory [16].

Given the machinery already provided by argumentation, the simplest way to build in degrees of belief is to translate every proposition in the belief unit that the agent is initially supplied with (which may contain nested modalities and so be of the form $B(I(\varphi))$) into an argument with an empty set of grounds. Thus $B(I(\varphi))$ becomes the argument:

$$(B(I(\varphi)) : \{\} : \alpha)$$

where α is the associated degree of belief. Any propositions which the agent gathers from external sources, such as its sensors or other agents, are entered as arguments with grounds which reflect their origins, for example:

$$(B(B(\varphi)) : \{\mathit{Other\ Agent}\} : \alpha)$$

Any propositions deduced from this base set will then accumulate grounds as detailed above, and the degree of belief of these deductions will depend upon the degrees of belief of its grounds. While we could use any of the many formalisms for handling uncertainty as a means of expressing this belief, in this paper we consider the degree of belief to be a mass assignment in the sense of the Dempster-Shafer theory [21].

Using this scheme, therefore, an agent's belief unit (which is its model of the world) is made up of a set of statements of the form:

$$\begin{array}{ll} 1 & (B(I(\varphi)) : \{\} : \alpha) \\ 2 & (B(I(\varphi)) \rightarrow B(I(\psi)) : \{\} : \beta) \end{array}$$

From these the agent can draw whatever deductions are sanctioned by the rules of inference of the logic of its belief unit. Thus, if the belief unit contains the rule of modus ponens, the agent can infer:

$$(B(I(\psi), \langle\{B(I(\varphi), B(I(\varphi)) \rightarrow B(I(\psi)\} \vdash_{\mathrm{agent}_{mp}} B(I(\psi)\rangle : \gamma)$$

where γ is the combination of α and β according to the Dempster-Shafer theory. When facts in the the belief base originate from external sources, the beliefs attached to those facts reflect the agent's view of the reliability of those sources [4].

There are a number of advantages in adding degrees of belief to the BDI framework. One we have mentioned before is the ability to model the world at a greater level of detail. Another is the possibility of increasing the granularity of the decision making process from one which in some situations picks an intention at random to one which picks an intention based upon its expected utility. A third, and the one we investigate here, is that of using the degrees of belief to inform the way in which the agent updates its model of the world when it learns new things.

5 Belief revision and updating

Both belief revision and updating are two different approaches which allow an agent to cope with a changing world by allowing it to alter its beliefs in response to new, possibly contradictory, information. The difference between the two actions is captured by the ideas of recoverability and priority of incoming information. Updating takes place when it is assumed that incoming information should take priority. It thus applies in situations in which the agent is operating in a dynamic world where the latest information is most correct. Belief revision, on the other hand, is appropriate when the agent is operating in a world for which it has only an approximate, incomplete or erroneous representation. In such cases the agent wants to take account of all available information. Belief revision consists of redefining the degrees of credibility of propositions in the light of incoming information. The model thus adopts the recoverability principle that any previously believed information item must belong to the current cognitive state if it is consistent with it. Unlike the case in which incoming information is given priority, this principle makes sure that the chronological sequence of the incoming information has nothing to do with the credibility of that information, and that the changes are not irrevocable. The following example should clarify the difference between revision and updating.

Example 2 Temporarily ignoring the use of arguments, suppose the belief unit of an agent contains the propositions $B(\alpha)$ and $B(\alpha) \rightarrow B(\beta)$. If the new proposition $\neg B(\beta)$ is observed we will have a contradiction between $B(\alpha), B(\alpha) \rightarrow B(\beta)$ and $\neg B(\beta)$ and consequently we will have three different consistent sets of facts:

1. $\{B(\alpha), \neg B(\beta)\}$
2. $\{\neg B(\beta), B(\alpha) \rightarrow B(\beta)\}$
3. $\{B(\alpha), B(\alpha) \rightarrow B(\beta)\}$

Using belief revision we can choose any one of these as a preferred maximally consistent set of beliefs, while when updating we can't choose the third because it doesn't contain the new information. $\square$

The model for belief revision we adopt is drawn from [4]. We start with the following definitions.

Definition 3 *The set of all facts that the agent is initially programmed with, its observations, and everything it is told by other agents is known as the set of* assumptions.

Definition 4 *The set of deductions that an agent can make from its set of assumptions are known as the* consequences.

Definition 5 *A* knowledge base *is the set of the assumptions introduced from the various sources, and a* knowledge space *is the set of all beliefs.*

The knowledge space (KS) is thus made up of the knowledge base (KB) plus all the deductions made from the knowledge base. Both the knowledge base and the knowledge space grow monotonically since none of their elements are ever erased from memory. Normally both contain *contradictions*.

Definition 6 *A* nogood *is defined as minimal inconsistent subset of a knowledge base. Dually, a* good *is a maximally consistent subset of a knowledge base.*

Thus a nogood is a subset of a KB that supports a contradiction and is not a superset of any other nogood. A good is a subset of a KB that is neither a superset of any nogood nor a subset of any other good. Each good has a corresponding *support set*, which is the subset of a KS made up of all the propositions that are in the good or are consequences of them. These definitions originate from de Kleer's work on assumption-based truth maintenance systems [3]. Procedurally, the method of belief revision consists of four steps:

S1 Generate the set NG of all the nogoods and the set G of all goods in the KB.

S2 Define a credibility ordering over the assumptions in the KB.

S3 Extend this into a credibility ordering over the goods in G.

S4 Select the preferred good CG with its corresponding support set SS.

The first step S1 deals with consistency and adopts the set-covering algorithm [20] to find NG and the corresponding G. S2 deals with uncertainty in the information. Since we have adopted the Dempster-Shafer theory of evidence [21] to handle uncertainty, S2 uses Dempster's rule of combination to find the credibility of the beliefs as above. In addition, S2 deals with any necessary revisions to the reliability of the sources of data using Bayesian conditioning (see [5] for details). This allows, for instance, the reliability of a source which conflicts with many others to be adjusted downwards. S3 also deals with uncertainty, but at the level of the goods, extending the ordering defined by S2 over the assumptions, into an ordering onto the goods. There are a number of possible methods for doing this [1], including *best-out*, *inclusion-based* and *lexicographic*. An alternative is to order the goods according to the average credibility of their elements. Doing this, however, means that the preferred good may no longer necessarily contain the most credible piece of information. Finally S4 consists of two substeps: selecting a good CG from G (normally, CG is the good with the highest credibility) and selecting from KS the derived sentences that are consequences of the propositions belong to CG.

If instead of revision, we require updating of beliefs there is no difference in the dynamics of the propagation of weights, the most of the precedure for calculating the new set of beliefs is the same. However, step S4 has to be modified so that incoming information replaces the old. As a result we are constrained to pick the preferred good from those which contain the new information. This gives the following procedure:

S1 Generate the set NG of all the nogoods and the set G of all goods in the KB.

S2 Define a credibility ordering over the assumptions in the KB.

S3 Extend this into a credibility ordering over the goods in G.

S4′ Select the preferred good CG which contains the new proposition, with its corresponding support set SS.

Recapitulating, for both revision and updating, we have:

INPUT:

- New proposition p;
- *KB*: set of all propositions introduced from the various sources; and
- Reliability of all sources.

OUTPUT:

- New credibilities of the propositions in $KB \cup \{p\}$;
- New credibilities of the goods in G;
- Preferred good CG and corresponding support set SS; and
- New reliability of all the sources.

Whether by revision or updating, the upshot of the procedure is that the beliefs of all assumptions, along with their sources, and the consequences of those assumptions are suitably altered. The following section contains an extended example of a belief revising agent which illustrates the way in which this happens.

6 An example

As an example of the use of the degrees of belief in the multi-context BDI model, let us consider an agent, Nico, who is investigating a murder. Nico has the following facts in her belief context:

1 $(dead(paolo) : \{\} : 1)$
2 $(shot(X, Y) \wedge dead(Y) \rightarrow murderer(Y, X) : \{\} : 1)$
3 $(was\_with(X, Y) \wedge murderer(Z, Y) \rightarrow suspected(X) : \{\} : 1)$
4 $(shot(benito, paolo) : \{Carl\} : 0.5)$

Thus Nico knows that Paolo is dead, and has some background information about murderers and their accomplices. All this information is certain. Nico also has information from a witness Carl which suggests that Benito shot Paolo, though Nico only judges Carl to be reliable to degree 0.5. From this, Nico can conclude that Benito murdered Paolo. This conclusion takes the form of the argument:

$$(murderer(paolo, benito) : \{1, 2, 4\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} murderer(paolo, benito)\rangle : 0.5)$$

where,

1. $murderer(paolo, benito)$ is the formulae which is the subject of the argument;
2. the terms $\{1, 2, 4\}$[2] are the grounds of the argument which may be used along with modus ponens—signified by the "mp"— by Nico to infer the fact that $murderer(paolo, benito)$; and

[2]These denote the formulae $dead(paolo)$, $shot(X, Y) \wedge dead(Y) \rightarrow murderer(Y, X)$ and $shot(benito, paolo)$ respectively.

3. 0.5 is the sign.

If new information that Ana was with Benito at the time of the shooting comes from a second witness Dana, whose reliability is 0.6, then Nico's belief context becomes:

1 $(dead(paolo) : \{\} : 1)$
2 $(shot(X,Y) \wedge dead(Y) \rightarrow murderer(Y,X) : \{\} : 1)$
3 $(was\_with(X,Y) \wedge murderer(Z,Y) \rightarrow suspected(X) : \{\} : 1)$
4 $(shot(benito, paolo) : \{Carl\} : 0.5)$
5 $(was\_with(ana, benito) : \{Dana\} : 0.6)$

Because Nico has some information about co-location and accomplicehood, Ana becomes a suspect in the killing, though Benito remains the main suspect:

$$(murderer(paolo, benito) : \langle \{1,2,4\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} murderer(paolo, benito) \rangle : 0.5)$$
$$(suspected(ana) : \langle \aleph, \{3,5,\Im\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} suspected(ana) \rangle : 0.3)$$

where $\Im$ denotes $murderer(paolo, benito)$ and $\aleph$ denotes the step:

$$\{1,2,4\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} murderer(paolo, benito)$$

Suppose now that a new information comes from the witness Dana that Benito did not shoot Paolo. This information is not compatible with the Nico's proposition number 5, so the belief revision process calculates new degrees of credibility for her beliefs and new reliabilities for Carl and Dana. After this process Nico's new belief context is:

1 $(dead(paolo) : \{\} : 1)$
2 $(shot(X,Y) \wedge dead(Y) \rightarrow murderer(Y,X) : \{\} : 1)$
3 $(was\_with(X,Y) \wedge murderer(Z,Y) \rightarrow suspected(X) : \{\} : 1)$
4 $(shot(benito, paolo) : \{Carl\} : 0.29)$
5 $(was\_with(ana, benito) : \{Dana\} : 0.42)$
6 $(\neg shot(benito, paolo) : \{Dana\} : 0.42)$

Now, consider that in the course of her investigations Nico finds further new evidence against Benito emerges when another agent, Ewan, whose reliability Nico judges be 0.9, says that Benito did shoot Paolo. When this happens, Nico's belief context changes once again—the belief revision mechanism starts from the reliabilities fixed *a priori* and Nico gets the following context:

1 $(dead(paolo) : \{\} : 1)$
2 $(shot(X,Y) \wedge dead(Y) \rightarrow murderer(Y,X) : \{\} : 1)$
3 $(was\_with(X,Y) \wedge murderer(Z,Y) \rightarrow suspected(X) : \{\} : 1)$
4 $(shot(benito, paolo) : \{Carl, Ewan\} : 0.88)$
5 $(was\_with(ana, benito) : \{Dana\} : 0.06)$
6 $(\neg shot(benito, paolo) : \{Dana\} : 0.06)$

From these facts she can build the following arguments:

$$(murderer(paolo, benito) : \langle \{1,2,4\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} murderer(paolo, benito) \rangle : 0.88)$$
$$(suspected(ana) : \langle \aleph, \{3,5,\Im\} \vdash_{\mathrm{Nico}_{\{\mathrm{mp}\}}} suspected(ana) \rangle : 0.06)$$

The result of all these revisions is that Nico is fairly sure that Carl and Ewan are reliable (as evidenced by the high belief given to the facts they supplied) and that Benito murdered Paolo. In addition, she believes that Dana is rather unreliable and so does not have much confidence that Ana is a suspect.

7 Summary

This paper has suggested a way of refining the treatment of beliefs in BDI models, in particular those built using multi-context systems as suggested in [14]. We believe that this work brings significant advantages. Firstly because the treatment is based upon the general ideas of argumentation, the approach we take is very general; it would, for instance, be simple to devise an analogous approach which made use of possibility measures rather than measures based on Dempster-Shafer theory. Secondly, the use of degrees of belief, as we have demonstrated, gives a plausible means of carrying out belief revision to handle inconsistent data, something that would be much harder to do in more conventional BDI models. Thirdly, introducing degrees of belief in propositions provides the foundation for using decision theoretic methods within BDI models; currently a topic which has had little attention. However, we acknowledge that this work is rather preliminary. In particular we need to extend the approach to deal with degrees of desire and intention, and to test out the approach in real applications. Both these directions are the topic of ongoing work.

References

[1] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 640–645, 1995.

[2] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, UK, 1980.

[3] J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[4] A. Dragoni and P. Giorgini. Belief revision through the belief function formalism in a multi-agent environment. In *Proceedings of the 3rd International Workshop on Agent Theories, Architectures and Languages*, 1996.

[5] A. Dragoni and P. Giorgini. Learning agents' reliability through Bayesian conditioning: a simulation study. In *Learning in DAI Systems*, pages 151–167, 1997.

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[7] J. Fox, P. Krause, and S. Ambler. Arguments, contradictions and practical reasoning. In *Proceedings of the 10th European Conference on Artificial Intelligence*, pages 623–627, 1992.

[8] J. Fox and S. Parsons. Arguments about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*. Springer Verlag, Berlin, 1998.

[9] D. Gabbay. *Labelled Deductive Systems*. Oxford University Press, Oxford, UK, 1996.

[10] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: How we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994.

[11] P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113–131, 1995.

[12] R. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.

[13] P. Noriega and C. Sierra. Towards layered dialogical agents. In *Proceedings of the 3rd International Workshop on Agents Theories, Architectures and Languages*, pages 157–171, 1996.

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

[15] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[16] A. Rao and M. Georgeff. Deliberation and its role in the formation of intentions. In *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, pages 300–307, 1991.

[17] A. Rao and M. Georgeff. BDI agents: From theory to practice. In *Proceedings of the 1st International Conference on Multi-Agent Systems*, pages 312–319, 1995.

[18] A. S. Rao and M. P. Georgeff. Modeling Rational Agents within a BDI-Architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484, 1991.

[19] A. S. Rao and M. P. Georgeff. Formal Models and Decision Procedures for Multi-Agent Systems. Technical Note 61, Australian Artificial Intelligence Institute, 1995.

[20] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 53, 1987.

[21] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

[22] F. H. van Eemeren, R. Grootendorst, F. S. Henkemans, J. A. Blair, R. H. Johnson, E. C. W. Krabbe, C. Plantin, D. N. Walton, C. A. Willard, J. Woods, and D. Zarefsky. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.