Agent specification using multi-context systems

Simon Parsons[1], Nicholas R. Jennings[2], Jordi Sabater[3], Carles Sierra[3]

[1] Department of Computer Science, University of Liverpool
Chadwick Building, Liverpool L69 7ZF, UK.[†]
`S.D.Parsons@csc.liv.ac.uk`
[2] Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton SO17 1BJ, UK
`nrj@ecs.soton.ac.uk`
[3] IIIA - Artificial Intelligence Research Institute
CSIC - Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia, Spain.
`{jsabater,sierra}@iiia.csic.es`

Abstract. In the area of agent-based computing there are many proposals for specific system architectures, and a number of proposals for general approaches to building agents. As yet, however, there are comparatively few attempts to relate these together, and even fewer attempts to provide methodologies which relate designs to architectures and then to executable agents. This paper discusses an attempt we have made to address this shortcoming, describing a general method of defining architectures for logic-based agents which can be directly executed. Our approach is based upon the use of multi-context systems and we illustrate its use through examples of the specification of a simple agents.

1 Introduction

Agent-based computing is fast emerging as a new paradigm for engineering complex, distributed systems [18, 36]. An important aspect of this trend is the use of agent architectures as a means of delivering agent-based functionality (as opposed to work on agent programming languages [19, 31, 34]). In this context, an architecture can be viewed as a separation of concerns—it identifies the main functions that ultimately give rise to the agent's behaviour and defines the interdependencies that exist between them. As agent architectures become more widely used, there is an increasing demand for unambiguous specifications of them and there is a greater need to verify implementations of them. To this end, a range of techniques have been used to formally specify agent architectures (including Concurrent MetateM [12, 35], DESIRE [3, 32] and Z [8]). However, these techniques typically fall short in at least one of the following ways: (i) they enforce a particular view of architecture upon the specification; (ii) they offer no explicit structures for modelling the components of an architecture or the relationships between

[†] Temporarily with the Center for Coordination Science, Sloan School of Management, Massachusetts Institute of Technology, 3 Cambridge Center NE20-336, Cambridge, MA 02142-1347, USA.

them; (iii) they leave a gap between the specification of an architecture and its implementation.

To rectify these shortcomings, we have proposed the use of *multi-context systems* [15] as a means of specifying and implementing agent architectures. Multi-context systems provide an overarching framework that allows distinct theoretical components to be defined and interrelated. Such systems consist of a set of contexts, each of which can informally be considered to be a logic and a set of formulae written in that logic, and a set of bridge rules for transferring information between contexts. Thus, different contexts can be used to represent different components of the architecture and the interactions between these components can be specified by means of the bridge rules between the contexts. We believe multi-context systems are well suited to specifying and modelling agent architectures for two main types of reason: (i) from a *software engineering perspective* they support modular decomposition and encapsulation; and (ii) from a *logical modelling perspective* they provide an efficient means of specifying and executing complex logics.

From a software engineering perspective, multi-context systems support the development of modular architectures. Each architectural component—be it a functional component (responsible for assessing the agent's current situation, say) or a data structure component (the agent's beliefs, say)—can be represented as a separate context. The links between the components can then be made explicit by writing bridge rules to link the contexts. This ability to directly support component decomposition offers a clean route from the high level specification of the architecture through to its detailed design. Moreover, this basic philosophy can be applied no matter how the architectural components are decomposed or how many architectural components exist.

Moving onto the logical modelling perspective, there are four main advantages of adopting a multi-context approach. The first is an extension of the software engineering advantages which specifically applies to logical systems. By breaking the logical description of an agent into a set of contexts, each of which holds a set of related formulae, we effectively get a form of many-sorted logic (all the formulae in one context are a single sort) with the concomitant advantages of scalability and efficiency. The second advantage follows on from this. 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 together 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 is illustrated below where we use multi-context systems to simplify the construction of a BDI agent.

Both of the above advantages apply to any logical agent built using multi-context systems. The remaining two advantages apply to specific types of logical agent—those which reason about their beliefs and those of other agents. The first is that multi-context systems make it possible [15] 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) but which obviates the difficulties usually inherent in theorem proving in such logics. Again this is illustrated in [23]. Thus the use of multi-context systems makes it easy to directly execute agent specifications where those specifications deal with modal notions. The final advantage is related to this. Agents which reason about beliefs are often confronted with the problem of modelling 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 [1, 6].

When the software engineering and the logical modelling perspectives are combined, it can be seen that the multi-context approach offers a clear path from specification through to implementation. By providing a clear set of mappings from concept to design, and from design to implementation, the multi-context approach offers a way of tackling the gap that currently exists between the theory and the practice of agent-based systems.

2 Multi-context agents

As discussed above, we believe that the use of multi-context systems offers a number of advantages when engineering agent architectures. However, multi-context systems are not a panacea. We believe that they are most appropriate when building agents which are logic-based and are therefore largely deliberative [1].

2.1 The basic model

Using a multi-context approach, an agent architecture consists of four basic types of component. These components were first identified in the context of building theorem provers for modal logic [15], before being identified as a methodology for constructing agent architectures [20]. The components are [2] :

– *Units*: Structural entities representing the main components of the architecture.
– *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.

Units represent the various components of the architecture. They contain the bulk of an agent's problem solving knowledge, and this knowledge is encoded in the specific theory that the unit encapsulates. In general, the nature of the units will vary between architectures. For example, a BDI agent may have units which represent theories of beliefs, desires and intentions (see Section 3), whereas an architecture based on a functional separation of concerns may have units which encode theories of cooperation, situation assessment and plan execution (see Section 4). In either case, each unit has a suitable logic associated with it. Thus the belief unit of a BDI agent has a logic of belief associated with it, and the intention unit has a logic of intention. The logic associated with each unit provides the language in which the information in that unit is encoded,

[1] See [38] for a discussion of the relative merits of logic-based and non logic-based approaches to specifying and building agent architectures.
[2] For more detail see [20].

and the bridge rules provide the mechanism by which information is transferred between units.

Bridge rules can be understood as rules of inference with premises and conclusions in different units. For instance:

$$\frac{u_1 : \psi, u_2 : \varphi}{u_3 : \theta}$$

means that formula θ may be deduced in unit u_3 if formulae ψ and φ are deduced in units u_1 and u_2 respectively.

When used as a means of specifying agent architectures, all the elements of the model, both units and bridge rules, are taken to work concurrently. In practice this means that the execution of each unit is a non-terminating, deductive process[3]. The bridge rules continuously examine the theories of the units 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 control elements.

2.2 The extended model

The model as outlined above is that introduced in [20] and used in [23]. However, this model has proved deficient in a couple of ways, both connected to the dynamics of reasoning. In particular we found it useful [29] to extend the basic idea of multi-context systems by associating two control elements with the bridge rules: *consumption* and *time-outs*. A consuming condition means the bridge rule removes the formula from the theory which contains the premise (remember that a theory is considered to be a set of formulae). Thus in bridge rules with consuming conditions, formulae "move" between units. To distinguish between a consuming condition and a non-consuming condition, we will use the notation $u_i > \psi$ for consuming and $u_i : \psi$ for non-consuming conditions. Thus:

$$\frac{u_1 > \psi, u_2 : \varphi}{u_3 : \theta}$$

means that when the bridge rule is executed, ψ is removed from u_1 but φ is not removed from u_2.

Consuming conditions increase expressiveness in the communication between units. With this facility, we can model the movement of a formula from one theory to another (from one unit to another), changes in the theory of one unit that cause the removal of a formula from another one, and so on. This mechanism also makes it possible to model the concept of state since having a concrete formula in one unit or another might represent a different agent state. For example, later in the paper we use the presence of a formula in a particular unit to indicate the availability of a resource.

A time-out in a bridge rule means there is a delay between the instant in time at which the conditions of the bridge rule are satisfied and the effective activation of the rule. A time-out is denoted by a label on the right of the rule; for instance:

$$\frac{u_1 : \psi}{u_2 : \varphi}[t]$$

[3] For more detail on exactly how this is achieved, see [29].

means that t units of time after the theory in unit u_1 gets formula ψ, the theory in unit u_2 will be extended by formula φ. If during this time period formula ψ is removed from the theory in unit u_1, this rule will not be applied. In a similar way to consuming conditions, time-outs increase expressiveness in the communication between units. This is important when actions performed by bridge rules need to be retracted if a specific event does not happen after a given period of time. In particular, it enables us to represent situations where silence during a period of time may mean failure (in this case the bridge rules can then be used to re-establish a previous state)[4].

2.3 Modular agents

Using units and bridge rules as the only structural elements can be cumbersome when building complex agents (as can be seen from the model we develop below in Section 3). As the complexity of the agent increases, it rapidly becomes very difficult to deal with the necessary number of units and their interconnections using bridge rules alone. Adding new capabilities to the agent becomes a complex task in itself. To solve this problem we suggest adding another level of abstraction to the model—the *module*.

A module is a set of units and bridge rules that together model a particular capability or facet of an agent. For example, planning agents must be capable of managing resources, and such an agent might have a module modeling this ability. Similarly, such an agent might have a module for generating plans, a module for handling communication, and so on. Thus modules capture exactly the same idea as the "capabilities" discussed by Busetta *et al.* [4]. Unlike Busetta *et al.*, we do not currently allow modules to be nested inside one another, largely because we have not yet found it necessary to do so. However, it seems likely that we will need to develop a means of handling nested hierachies of modules in order to build more complex agents than we are currently constructing.

Each module must have a communication unit. This unit is the module's unique point of contact with the other modules and it knows what kind of messages its module can deal with. All of an agent's communication units are inter-connected with the others using *multicast bridge rules* (*MBR*s) as in Figure 1. This figure shows three MBRs (the rectangles in the middle of the diagram) each of which has a single premise in module a and a single conclusion in each of the modules $\mathsf{n_i}$.

Since the MBRs send messages to more than one module, a single message can provoke more than one answer and, hence, contradictory information may appear. There are many possible ways of dealing with this problem, however here we consider just one of them as an example. We associate a weight with each message. This value is assigned to the message by the communication unit of the module that sends it out. Weights belong to $[0, 1]$ (maximum importance is 1 and minimum is 0), and their meaning is the

[4] Both of these extensions to the standard multi-context system incur a cost. This is that including them in the model means that the model departs somewhat from first order predicate calculus, and so does not have a fully-defined semantics. We are currently looking at using linear logic, in which individual propositions can only be used once in any given proof, as a means of giving a semantics to consuming conditions, and various temporal logics as a means of giving a semantics to time-outs.

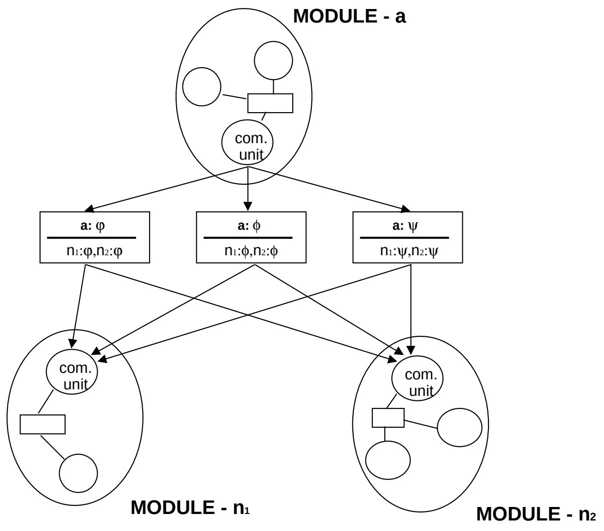

Fig. 1. The inter-connection of modules (from a's perspective only)

strength of the opinion given in the message, and this can be used to resolve contradictory messages. For instance, the message with highest weight might be preferred, or the different weights of incoming messages could be combined by a communication unit receiving them to take a final decision (for instance using the belief revision mechanism described in [21]). Note that weights are used only in *inter-module* messages.

2.4 Messages between modules

Given a set AN of agent names and a set MN of module names, an inter-module message has the form:

$$I(S, R, \varphi, G, \psi)$$

where

- I is an illocutionary particle that specifies the kind of message.
- S and R both have the form $A[/m]^*$[5] where $A \in AN$ or $A = \mathit{Self}$ (Self refers to the agent that owns the module) and $m \in MN$, or $m = \mathit{all}$ (all denotes all the modules within that agent). S reflects who is sending the message and R indicates to whom it is directed.
- φ is the content of the message.

[5] As elsewhere we use BNF syntax, so that $A[/m]^*$ means A followed by one or more occurrences of $/m$.

- G is a record of the derivation of φ. It has the form: $\{\{\Gamma_1 \vdash \varphi_1\} \ldots \{\Gamma_n \vdash \varphi_n\}\}$ where Γ is a set of formulae and φ_i is a formula with $\varphi_n = \varphi$ [6].
- $\psi \in [0, 1]$ is the weight associated with the message.

To see how this works in practice, consider the following. Suppose that an agent (named B) has four modules (*a*, *b*, *c*, *d*). Module *a* sends the message:

$$Ask(Self/a, Self/all, Give(B, A, Nail), \psi_1, 0.5)$$

This means that module a of agent B is asking all its modules whether B should give A a nail. The reason for doing this is ψ_1 and the weight a puts on this request is 0.5. Assume modules *c* and *d* send the answer

$$Answer(Self/c, Self/a, not(Give(B, A, Nail)), \psi_2, 0.6)$$

and

$$Answer(Self/d, Self/a, not(Give(B, A, Nail)), \psi_3, 0.7)$$

while module *b* sends

$$Answer(Self/b, Self/a, Give(B, A, Nail), \psi_4, 0.3)$$

Currently we treat the weights of the messages as possibility measures [9], and so combine the disjunctive support for $not(Give(B, A, Nail))$ using max. As this combined weight is higher than the weight of the positive literal, the communication unit of module *a* will accept the opinion $not(Give(B, A, Nail))$.

The messages we have discussed so far are those which are passed around the agent itself in order to exchange information between the modules which compose it. Our approach also admits the more common idea of messages between agents. Such inter-agent messages have the same basic form, but they have two minor differences:

- S and R are agent names (i.e. $S, R \in AN$), no modules are specified.
- there is no degree of importance (because it is internal to a particular agent—however inter-agent messages could be augmented with a degree of belief [21] which could be based upon the weight of the relevant intra-agent messages.)

With this machinery in place, we are in a position to specify realistic agent architectures.

2.5 Examples of multi-context agents

This remainder of this paper contains two examples of agent specification using multi-context systems, each illustrating one of the uses of units introduced in Section **??**—the first of these (based on the model in [23]) is that for a BDI agent, the second (based on

[6] In other words, G is exactly the set of grounds of the argument for φ [23]. Where the agent does not need to be able to justify its statements, this component of the message can be discarded. Note that, as argued by Gabbay [13] this approach is a generalisation of classical logic—there is nothing to stop the same approach being used when messages are just formulae in classical logic.

the model in [29]), is that for an agent in which the architectural units are based on a functional separation of concerns. The first illustrates how the multi-context approach can be used to handle the kind of “mental attitudes” agent architectures which have become common. The second shows how modules can help to simplify the multi-context model.

Both of these examples are based around the example of home improvement agents introduced in [22], and sketched below[7]. In order to save space (and also to save the sanity of the authors and readers familiar with the example), neither treatment does any more than specify the agents—fuller versions can be found in the papers cited above.

For those unfamiliar with the example, it is as follows. Two agents, A and B have, respectively, the tasks of hanging a picture and hanging a mirror. A knows one way of hanging a picture and one of hanging a mirror. B just knows how to hang a mirror (using a different technique from A). A has the means to hang a mirror using its technique, B has the means to hand either its mirror, using its own technique, or A’s picture. The full solution to the problem involves A convincing B to use A’s approach and resources to hang the mirror so that A can use B’s resources to hang the picture.

3 Agents with mental attitudes

Our first example examines how a particular class of agent architecture—BDI agents—can be modelled and then describes how particular individuals of that class can be specified in order to solve the example. This seems an appropriate choice because BDI agents are currently of wide interest within the multi-agent system community [37].

3.1 A high-level description

The first step in specifying the agent is to choose the units and the logics that they contain. In this example, the choice is driven by the fact that we are modelling BDI agents. The particular theory of BDI on which the architecture is based is that of Rao and Georgeff. This model has evolved over time (as can be seen by comparing [25] and [26]) and in this section we account for the most recent approach [26] where three modalities are distinguished: B for beliefs—used to represent the state of the environment, D for desires—used to represent the motivations of the agent, and I for intentions—used to represent the ends (or goals) of the agent. In order to fit this kind of model into our multi-context framework, we associate a separate unit for each of the modalities[8]

As dicussed in [23], we could then equip each of these units with exactly the same logic as is used in Rao and Georgeff’s model, taking the logic of the belief unit to be modal logic KD45 and the logics of the desire and intention units to both be modal logic KD, and to take all these modal logics to be combined with the temporal logic CTL [10]. However, it is more in the spirit of multi-context systems [15] to take B, D and I as predicates. Such systems again have separate B, D and I units along with

[7] Initially unnamed, this example seems to have become known as “The Nail Problem” (TNP); despite being simple to express it turns out to be rather hard to handle.

[8] In fact the general approach allows more than one unit for beliefs (as in [5]), desires or intentions if deemed appropriate. In the examples presented, however, this is not necessary.

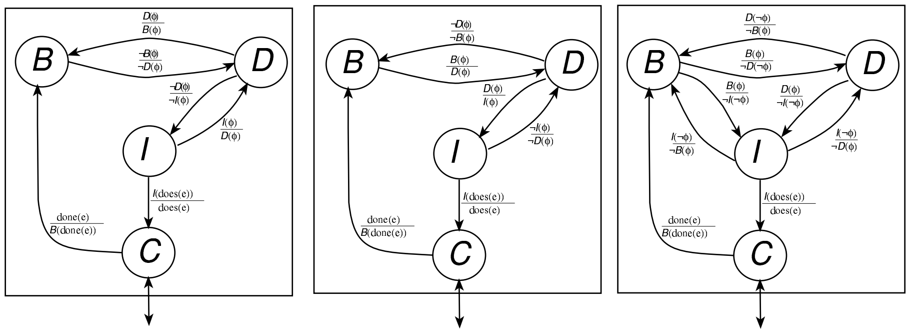

Fig. 2. Different types of BDI agent. From left to right, the relations between modalities correspond to strong realism, realism and weak realism.

a communication unit, and use first order logic. The necessary interaction between the predicates is established using bridge rules (as discussed below) and the axioms of the relevant modal logics are modelled by adding formulae to the theories in each unit (again this is discussed below).

3.2 Specification of bridge rules

Having decided on the units and the logics that they contain, the next step in the specification is to write down the bridge rules which connect the units. Here we have two distinct sets of such rules. The first model the relationships between beliefs, desires and intentions. These are domain independent and would hold for any BDI agent specified in this way. The second model some domain specific knowledge.

BDI bridge rules As stated above, the set of bridge rules determine the relationship between the modalities and hence the behaviour of the agent. Three well established sets of relationships for BDI agents have been identified [26]:

- *Strong realism*. 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 nor intend it [25].
- *Realism*. The set of beliefs is a subset of the set of desires which in turn is a subset of the set of intentions. That is, if an agent believes something, it both desires and intends it [7].
- *Weak realism*. A case in between strong realism and realism. Agents do not desire properties the negation of which are believed, do not intend propositions the negations of which are desired, and do not intend propositions the negations of which are believed [24].

Figure 2 gives a suitable set of bridge rules for each of these interpretations. In [23], we only considered strong realist agents. In addition to this set of rules, we found that we needed a couple of additional rules which relate intentions to beliefs:

$$\mathsf{AWARENESS\_OF\_INTENTION}(1) = \frac{I : I_i(\alpha)}{B : B_i(I_i(\alpha))}$$
$$\mathsf{AWARENESS\_OF\_INTENTION}(2) = \frac{I : \neg I_i(\alpha)}{B : B_i(\neg I_i(\alpha))}$$

Agents are aware of their intentions, so if an agent has an intention it also believes that it has that intention.

$$\mathsf{IMPLUSIVENESS} = \frac{B : B_i(I_i(\alpha))}{I : I_i(\alpha)}$$

When an agent believes it has an intention, it adopts that intention.

These last two are similar in some ways to the basic rules of modal logic[9], except that in standard modal logic they don't apply across modalities in the way that they do here.

Domain dependent bridge rules The bridge rules for the appropriate form of realism will be required for the specification of any such agent whatever domain it is operating in. Without them, the agent will not conform to Rao and Georgeff's idea of what a BDI agent is. In addition we believe that the awareness of intentions and implusiveness rules (or something like them) will be required in practice by any BDI agent.

In addition to these domain independent rules, any agent will require a set of bridge rules which define how it interacts with other agents. In the domain of this example, these relate the mental state to what an agent says (and what it hears to its mental state). These are as follows[10]:

$$\mathsf{REQUEST} = \frac{I : I_i(Give(X, i, Z))}{C : Ask(i, X, Give(X, i, Z))}$$

When an agent (i) needs something (Z) from another agent (X), it asks for it

$$\mathsf{OFFER} = \frac{I : I_i(Give(i, X, Z))}{C : Tell(i, X, Give(i, X, Z))}$$

When an agent (i) has the intention of offering something (Z) to another agent (X), it informs the recipient of this fact.

$$\mathsf{TRUST} = \frac{C : Tell(X, i, B_X(\varphi))}{B : B_i(\varphi)}$$

When an agent (i) is told of a belief of another agent (X), it accepts that belief.

$$\mathsf{AWARENESS\_OF\_ILLOCUTIONS} = \frac{C : \alpha}{B : B_i(\alpha)}$$

[9] In particular the positive and negative introspection axioms 4 and 5 and the T axiom.

[10] Note that in the rest of the paper we adopt a Prolog-like notation in which the upper case letters X, Y, Z, P are taken to be variables.

In addition, Figure 2 includes some bridge rules which allow the transfer of information between the communication unit and the belief and intention units. These capture the fact that an agent with an intention to carry out an action will communicate that fact, and when an agent receives notification that another agent has carried out an action, the first agent believes this.

This completes the set of bridge rules that we require for our example, and we can pass on to consider the logical theories with which each unit is instantiated. However, before doing so, consider that we have now specified 13 bridge rules [11] to connect 4 units. It is this tight network of interconnection that led us to conside the modular approach described in Section 2.3.

3.3 Instantiating the contexts

Having specified the contexts, logics and bridge rules we have to consider what formulae will appear in each unit. Some of these will be specific to an individual agent (the desires with which it is programmed for example), but others will be more generic and be common between a number of agents. It is these more generic formulae that we consider here. In the case of the home improvement agents, both agents need a simple theory of action that integrates a model of the available resources with their planning mechanism. This theory needs to model the following ideas (where i is an index identifying the agent):

Ownership. When an agent (X) is the owner of an artifact (Z) and it gives Z to another agent (Y), Y becomes its new owner:

$$B : B_i(Have(X, Z) \wedge Give(X, Y, Z) \rightarrow Have(Y, Z))$$

Unicity. When an agent (X) gives an artifact (Z) away, it no longer owns it [12]:

$$B : B_i(Have(X, Z) \wedge Give(X, Y, Z) \rightarrow \neg Have(X, Z))$$

Benevolence. When an agent i has something (Z) that it does not intend to use and is asked to give it to another agent (X), i adopts the intention of giving Z to X. Naturally more complex cooperative strategies can be defined if desired:

$$B : B_i(Have(i, Z) \wedge \neg I_i(Have(i, Z)) \wedge Ask(X, i, Give(i, X, Z)) \rightarrow I_i(Give(i, X, Z)))$$

[11] That is we require 13 in order to specify a strong realist agent. A weak realist agent would require 15.

[12] As it stands this formula appears contradictory. This is because we have, for simplicity, ignored the treatment of time. Of course, the complete specification of this example (which is not our main focus) would need time to be handled. We could do this by including time as an additional argument to each predicate, in which case the unicity formula would read $B : B_i(Have(X, Z, t) \wedge Give(X, Y, Z, t) \rightarrow \neg Have(X, Z, t + 1))$. Doing this would involve making the base logic for each unit "time capable", for instance by using the system introduced by Vila [33].

The following axioms represent a similarly simplistic theory of planning (but again one which suffices for our example). In crude terms, when an agent believes that it has the intention of doing something and has a rule for achieving that intention then the pre-conditions of the rule become new intentions. Recall that the $\rightarrow$ between the P_i and Q is not material implication.

Parsimony. If an agent believes that it does not intend something, it does not believe that it will intend the means to achieve it.

$$B : B_i(\neg I_i(Q)) \wedge B_i(P_1 \wedge \ldots \wedge P_j \wedge \ldots \wedge P_n \rightarrow Q) \rightarrow \neg B_i(I_i(P_j))$$

Reduction. If there is only one way of achieving an intention, an agent adopts the intention of achieving its preconditions.

$$\begin{aligned} B : B_i(I_i(Q)) \wedge B_i(P_1 \wedge \ldots \wedge P_j \wedge \ldots \wedge P_n \rightarrow Q) \\ \wedge \neg B_i(R_1 \wedge \ldots \wedge R_m \rightarrow Q) \rightarrow B_i(I_i(P_j)) \end{aligned}$$

where $R_1 \wedge \ldots \wedge R_m$ is not a permutation of $P_1 \wedge \ldots \wedge P_n$.

Unique Choice. If there are two or more ways of achieving an intention, only one is intended. Note that we use $\bigtriangledown$ to denote exclusive or.

$$\begin{gathered} B : B_i(I_i(Q)) \wedge B_i(P_1 \wedge \ldots \wedge P_j \wedge \ldots \wedge P_n \rightarrow Q) \\ \wedge B_i(R_1 \wedge \ldots \wedge R_m \rightarrow Q) \rightarrow \\ B_i(I_i(P_1 \wedge \ldots \wedge P_n)) \bigtriangledown B_i(I_i(R_1 \wedge \ldots \wedge R_m)) \end{gathered}$$

where $R_1 \wedge \ldots \wedge R_m$ is not a permutation of $P_1 \wedge \ldots \wedge P_n$. As mentioned above, we acknowledge that both the theory of action and the theory of planning are rather naive. The interested reader is encouraged to substitute their own such theories if desired.

So far, we have identified the contexts and the logics they will contain, decided on the bridge rules between them, and identified the bits of the theories expressed in each logic that are common to both agents in our example. It remains to add to the model those bits of the theories that are unique to each agent.

3.4 Instantiating the individual agents

Agent a has the intention of hanging a picture, it has various beliefs about resources and how they can be used to hang mirrors and pictures:

$$\begin{aligned} & I : I_a(Can(a, hang\_picture)) \\ & B : B_a(Have(a, picture)) \\ & B : B_a(Have(a, screw)) \\ & B : B_a(Have(a, hammer)) \\ & B : B_a(Have(a, screwdriver)) \\ & B : B_a(Have(b, nail)) \\ & B : B_a(Have(X, hammer) \wedge Have(X, nail) \wedge Have(X, picture) \rightarrow \\ & \qquad Can(X, hang\_picture)) \\ & B : B_a(Have(X, screw) \wedge Have(X, screwdriver) \wedge Have(X, mirror) \rightarrow \\ & \qquad Can(X, hang\_mirror)) \end{aligned}$$

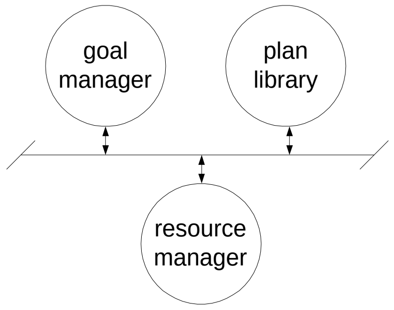

Fig. 3. The modules in the agent

Now, agent b wants to hang a mirror (and has this as an intention) and has various beliefs about its resources and the action of hanging mirrors:

$$
\begin{aligned}
I &: I_b(Can(b, hang\_mirror)) \\
B &: B_b(Have(b, mirror)) \\
B &: B_b(Have(b, nail)) \\
B &: B_b(Have(X, hammer) \wedge Have(X, nail) \wedge Have(X, mirror) \rightarrow \\
&\quad\; Can(X, hang\_mirror))
\end{aligned}
$$

We have now demonstrated how the multi-context approach can be used to specify BDI agents. As mentioned above, [23] shows how this specification can be used to solve the example.

4 A functional agent

This section gives a specification of an agent which is capable of solving a simplified version of the home-improvement example. The simplification is to reduce the problem to one in which a single agent has all the resources necessary to hang a picture. As a result, compared with the more complex versions of the home improvement agents described above, the agent is not quite solipsistic (since it has some awareness of its environment) but it is certainly autistic (since it has no mechanisms for interacting with other agents). For an example of the specification of further agents in the context of this example, see [29, 27][13].

4.1 A high-level description

The basic structure of the agent is that of Figure 3. There are three modules connected by multicast bridge rules. These are the plan library (PL), the resource manager (RM),

[13] Note that [27] is distinct from [28]. The former is the version in the workshop preproceedings, whereas the latter is the version available in the published proceedings and the examples they contain are substantially different.

and the goal manager (GM). Broadly speaking, the plan library stores plans for the tasks that the agent knows how to complete, the resource manager keeps track of the resources available to the agent, and the goal manager relates the goals of the agent to the selection of appropriate plans.

There are two types of message which get passed along the multicast bridge rules. These are the following:

- **Ask:** a request to another module.
- **Answer:** an answer to an inter-module request.

Thus all the modules can do is to make requests on one another and answer those requests. We also need to define the predicates which form the content of such messages. Given a set of agent names AN, and with $AN' = AN \cup \{Self\}$.

- $Goal(X)$: X is a string describing an action. This denotes the fact that the agent has the goal X.
- $Have(X, Z)$: $X \in AN'$ is the name of an agent (here always instantiated to $Self$, the agent's name for itself, but a variable since the agent is aware that other agents may own things), and Z is the name of an object. This denotes Agent X has possession of Z.

As can be seen from the above, the content of the messages is relatively simple, referring to goals that the agent has, and resources it possesses. Thus a typical message would be a request from the goal manager as to whether the agent possesses a hammer:

$$ask(Self/GM, Self/all, goal(have(Self, hammer)), \{\})$$

Note that in this message, as in all messages in the remainder of this paper, we ignore the weight in the interests of clarity. Such a request might be generated when the goal manager is trying to ascertain if the agent can fulfill a possible plan which involves using a hammer.

4.2 Specifications of the modules

Having identified the structure of the agent in terms of modules, the next stage in the specification is to detail the internal structure of the modules in terms of the units they contain, and the bridge rules connecting those units. The structure of the plan library module is given in Figure 4. In this diagram, units are represented as circles, and bridge rules as rectangles. Arrows into bridge rules indicate units which hold the antecedents of the bridge rules, and arrows out indicate the units which hold the consequents. The two units in the plan library module are:

- The communication unit (CU): the unit which handles communication with other units.
- The plan repository (S): a unit which holds a set of plans.

The bridge rule connecting these units is:

$$\mathsf{GET\_PLAN} = \frac{\begin{array}{c} CU > ask(Self/Sender, Self/all, goal(Z), \{\}), \\ S : plan(Z, P) \end{array}}{CU : answer(Self/PL, (Self/Sender, goal(Z), \{P\})}$$

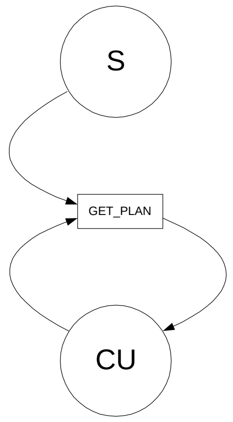

Fig. 4. The plan library module

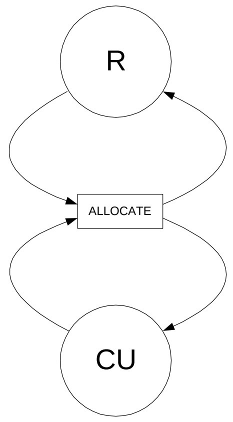

Fig. 5. The resource manager module

where the predicate $plan(Z, P)$ denotes the fact that P, taken to be a conjunction of terms, is a plan to achieve the goal Z[14].

When the communication unit sees a message on the inter-module bus asking about the feasibility of the agent achieving a goal, then, if there is a plan to achieve that goal in the plan repository, that plan is sent to the module which asked the original question. Note that the bridge rule has a consuming condition—this is to ensure that the question is only answered once.

The structure of the resource manager module is given in Figure 5. The two units in this module are:

– The communication unit (CU).

[14] Though here we take a rather relaxed view of what constitutes a plan—our "plans" are little more than a set of pre-conditions for achieving the goal.

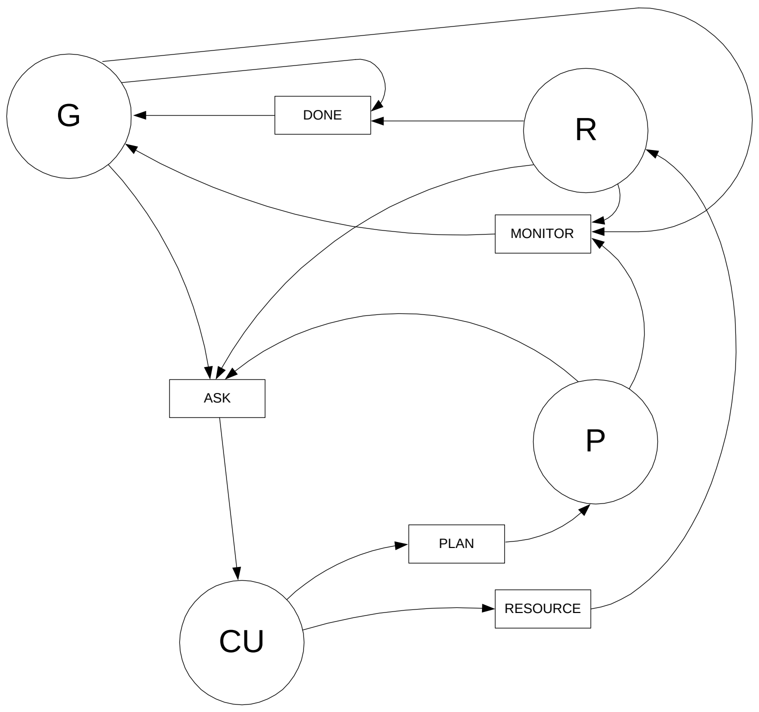

Fig. 6. The goal manager module

– The resource respository (R): a unit which holds the set of resources available to the agent.

The bridge rule connecting the two units is the following:

$$\mathsf{ALLOCATE} = \frac{\begin{array}{c} CU > ask(Self/Sender, Self/Receiver, goal(have(X, Z)), \{\}), \\ R > resource(Z, free) \end{array}}{\begin{array}{c} CU : answer(Self/RM, Self/Sender, have(X, Z), \{\}), \\ R : resource(Z, allocated) \end{array}}$$

where the $resource(Z, allocated)$ denotes the fact that the resource Z is in use, and $resource(Z, free)$ denotes the fact that the resource Z is not in use.

When the communication unit sees a message on the inter-module bus asking if the agent has a resource, then, if that resource is in the resource repository and is currently free, the formula recording the free resource is deleted by the consuming condition, a new formula recording the fact that the resource is allocated is written to the repository, and a response is posted on the inter-module bus. Note that designating a resource as “allocated” is not the same as consuming a resource (which would be denoted by the deletion of the resource), and that once again the bridge rule deletes the original message from the communication unit.

The goal manager is rather more complex than either of the previous modules we have discussed, as is immediately clear from Figure 6 which shows the modules it contains, and the bridge rules which connect them. These modules are:

– The communication unit (CU).
– The plan list unit (P): this contains a list of plans the execution of which is currently being monitored.
– The goal manager unit (G): this is the heart of the module, and ensures that the necessary sub-goaling is carried out.
– The resource list module (R): this contains a list of the resources being used as part of plans which are currently being executed.

The bridge rules relating these units are as follows. The first two bridge rules handle incoming information from the communication unit:

$$\mathsf{RESOURCE} = \frac{CU > answer(Self/RM, Self/GM, have(Self, Z), \{\})}{R : Z}$$
$$\mathsf{PLAN} = \frac{CU > answer(Self/PL, Self/GM, goal(Z), \{P\})}{P : plan(Z, P)}$$

The first of these, RESOURCE, looks for messages from the resource manager reporting that the agent has possession of some resource. When such a message arrives, the goal manager adds a formula representing the resource to its resource list module. The second bridge rule PLAN does much the same for messages from the plan library reporting the existence of a plan—such plans are written to the plan library. There is also a bridge rule ASK which generates messages for other modules:

$$\mathsf{ASK} = \frac{\begin{array}{c} G : goal(X), \\ G : not(done(X)), \\ R : not(X), \\ P : not(plan(X, Z)) \\ G : not(done(ask(X))), \end{array}}{\begin{array}{c} CU : ask(Self/G, Self/all, goal(X), \{\}), \\ G : done(ask(X)) \end{array}}$$

If the agent has the goal to achieve X, and X has not been achieved, nor is X an available resource (and therefore in the R unit), nor is there a plan to achieve X, and X has not already been requested from other modules, then X is requested from other modules and this request is recorded. The remaining bridge rules are:

$$\mathsf{MONITOR} = \frac{\begin{array}{c} G : goal(X), \\ R : not(X), \\ P : plan(X, P) \end{array}}{G : monitor(X, P)}$$
$$\mathsf{DONE} = \frac{\begin{array}{c} G : goal(X), \\ R : X \end{array}}{G : done(X)}$$

The MONITOR bridge rule takes a goal X and, if there is no resource to achieve X but there is a plan to obtain the resource, adds the formula $monitor(X, P)$ to the G unit, which has the effect of beginnning the search for the resources to carry out the plan. The DONE bridge rule identifies that a goal X has been achieved when a suitable resource has been allocated.

4.3 Specifications of the units

Having identified the individual units within each module, and the bridge rules which connect the units, the next stage of the specification is to identify the logics present within the various units, and the theories which are written in those logics. For this agent most of the units are simple containers for atomic formulae. In contrast, the G unit contains a theory which controls the execution of plans. The relevant formulae are:

$$\begin{aligned} monitor(X, P) &\rightarrow assert\_subgoals(P) \\ monitor(X, P) &\rightarrow prove(P) \\ monitor(X, P) \wedge proved(P) &\rightarrow done(X) \\[1em] assert\_subgoals(\bigwedge_i Y_i) &\rightarrow \bigwedge_i goal(Y_i) \\ prove(X \wedge \bigwedge_i Y_i) \wedge done(X) &\rightarrow prove(\bigwedge_i Y_i) \\ \bigwedge_i done(Y_i) &\rightarrow proved(\bigwedge_i Y_i) \end{aligned}$$

The $monitor$ predicate forces all the conjuncts which make up its first argument to be goals (which will be monitored in turn), and kicks off the "proof" of the plan which is its second argument[15]. This plan will be a conjunction of actions, and as each is "done" (a state of affairs achieved through the allocation of resources by other bridge rules), the proof of the next conjunct is sought. When all have been "proved", the relevant goal is marked as completed.

The specification as presented so far is generic—it is akin to a class description for a class of autistic home improvement agents. To get a specific agent we have to "program" it by giving it information about its initial state. For our particular example there is little such information, and we only need to add formulae to three units. The plan repository holds a plan for hanging pictures using hammers and nails:

$$\begin{aligned} &S : plan(hangPicture(X), \\ &\quad have(X, picture) \wedge have(X, nail) \wedge have(X, hammer)) \end{aligned}$$

The resource repository holds the information that the agent has a picture, nail and a hammer:

$$R : Resource(picture, free)$$

[15] Given our relaxed view of planning, this "proof" consists of showing the pre-conditions of the plan can be met.

$$R : Resource(nail, free)$$
$$R : Resource(hammer, free)$$

Finally, the goal manager contains the fact that the agent has the goal of hanging a picture:

$$G : goal(hangPicture(Self))$$

With this information, the specification is complete. A full description of the execution of this specification is contained in [28].

5 Related Work

There are two main strands of work to which ours is related—work on executable agent architectures and work on multi-context systems. As mentioned above, most previous work which has produced formal models of agent architectures, for example dMARS [16], Agent0 [30] and GRATE* [17], has failed to carry forward the clarity of the specification into the implementation—there is a leap of faith required between the two. Our work, on the other hand, maintains a clear link between specification and implementation through the direct execution of the specification as exemplified in our running example. This relation to direct execution also distinguishes our work from that on modelling agents in Z [8], since it is not yet possible to directly execute a Z specification. It is possible to animate specifications, which makes it possible to see what would happen if the specification were executed, but animating agent specifications is some way from providing operational agents. Our work also differs from that which aims to describe the operational semantics of agent architectures using the π-calculus [11], since our models have a declarative rather than an operational semantics.

More directly related to our work is that on DESIRE and Concurrent MetateM. DESIRE [3, 32] is a modelling framework originally conceived as a means of specifying complex knowledge-based systems. DESIRE views both the individual agents and the overall system as a compositional architecture. All functionality is designed as a series of interacting, task-based, hierarchically structured components. Though there are several differences, from the point of view of the proposal advocated in this paper, we can see DESIRE's *tasks* as modules and *information links* as bridge rules. In our approach there is no an explicit task control knowledge of the kind found in DESIRE. There are no entities that control which units, bridge rules or modules should be activated nor when and how they are activated. Also, in DESIRE the communication between tasks is carried out by the information links that are wired-in by the design engineer. Our inter-module communication is organized as a bus and the independence between modules means new ones can be added without modifying the existing structures. Finally the communication model in DESIRE is based on a one-to-one connection between *tasks*, in a similar way to that in which we connect units inside a module. In contrast, our communication between modules is based on a multicast model.

Concurrent MetateM defines concurrent semantics at the level of single rules [12, 35]. Thus an agent is basically a set of temporal rules which fire when their antecedents are satisfied. Our approach does not assume concurrency within the components of units, rather the units themselves are the concurrent components of our architectures. This means that our model has an inherent concurrent semantics at the level of the units and has no central control mechanism. Though our exemplar uses what is essentially first order logic (albeit a first order logic labelled with arguments), we could use any logic we choose—we are not restricted to a temporal logic as in MetateM.

There are also differences between our work and previous work on using multi-context systems to model agents' beliefs. In the latter [14], different units, all containing a belief predicate, are used to represent the beliefs of the agent and the beliefs of all the acquaintances of the agent. The nested beliefs of agents may lead to tree-like structures of such units (called *belief contexts*). Such structures have then been used to solve problems like the three wise men [6]. In our case, however, any nested beliefs would typically be included in a single unit or module. Moreover we provide a more comprehensive formalisation of an autonomous agent in that we additionally show how capabilities other than that of reasoning about beliefs can be incorporated into the architecture. In this latter respect this paper extends the work of [23] with the idea of modules which links the approach more strongly with the software engineering tradition.

6 Conclusions

This paper has proposed a general approach to defining agent architectures. It provides a means of structuring logical specifications of agents in a way which makes them directly executable. This approach has a number of advantages. Firstly it bridges the gap between the specification of agents and the programs which implement those specifications. Secondly, the modularity of the approach makes it easier to build agents which are capable of carrying out complex tasks such as distributed planning. From a software engineering point of view, the approach leads to architectures which are easily expandable, and have re-useable components.

From this latter point of view, our approach suggests a methodology for building agents which has similarities with object-oriented design [2]. The notion of inheritance can be applied to groups of units and bridge rules, modules and even complete agents. These elements could have a general design which is specialized to different and more concrete instances by adding units and modules, or by refining the theories inside the units of a generic agent template. However, before we can develop this methodology, there are some issues to resolve. Firstly there is the matter of the semantics of the comsuming conditions and time-outs in bridge rules. Secondly, there is the question of how to handle nested hierachies of modules—something which is essential if we are to develop really complex agents.

Acknowledgments This work has been supported by the UK EPSRC project Practical Negotiation for Electronic Commerce GR/M07076 and the EU IST project Sustainable Lifecycles for Information Ecosystems IST-1999-10208.

References

1. M. Benerecetti, A. Cimatti, E. Giunchiglia, F. Giunchiglia, and L. Serafini. Formal specification of beliefs in multi-agent systems. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III*, pages 117–130. Springer Verlag, Berlin, 1997.
2. G. Booch. *Object-oriented analysis and design with application*. Addison Wesley, Wokingham, UK, 1994.
3. F. M. T. Brazier, B. M. Dunin-Keplicz, N. R. Jennings, and J. Treur. Formal specification of multi-agent systems. In *Proceedings of the 1st International Conference on Multi-Agent Systems*, pages 25–32, 1995.
4. P. Busetta, N. Howden, R. Ronnquist, and A. Hodgson. Structuring BDI agents in functional clusters. In N. R. Jennings and Y Lespérance, editors, *Intelligent Agents VI*. Springer Verlag, Berlin, 1999.
5. A. Cimatti and L. Serafini. Multi-agent reasoning with belief contexts: The approach and a case study. In *Proceedings of the 3rd International Workshop on Agent Theories, Architectures and Languages*, 1994.
6. A. Cimatti and L. Serafini. Multi-agent reasoning with belief contexts: The approach and a case study. In M. J. Wooldridge and N. R. Jennings, editors, *Intelligent Agents*, pages 62–73. Springer Verlag, Berlin, 1995.
7. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
8. M. d'Inverno, D. Kinny, M. Luck, and M. Wooldridge. A formal specification of dMARS. In M. P. Singh, A. S. Rao, and M. Wooldridge, editors, *Intelligent Agents IV*, pages 155–176. Springer Verlag, Berlin, 1998.
9. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, NY, 1988.
10. E. A. Emerson. Temporal and Modal Logic. In J van Leeuwen, editor, *Handbook of Theoretical Computer Science*, pages 996–1071. Elsevier, 1990.
11. J. Ferber and O. Gutknecht. Operational semantics of a role-based agent architecture. In N. R. Jennings and Y Lespérance, editors, *Intelligent Agents VI*. Springer Verlag, Berlin, 1999.
12. M. Fisher. Representing abstract agent architectures. In J. P. Müller, M. P. Singh, and A. S. Rao, editors, *Intelligent Agents V*, pages 227–242. Springer Verlag, Berlin, 1998.
13. D. Gabbay. *Labelled Deductive Systems*. Oxford University Press, Oxford, UK, 1996.
14. F. Giunchiglia. Contextual reasoning. In *Proceedings of the IJCAI Workshop on Using Knowledge in Context*, 1993.
15. F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: How we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994.
16. F. F. Ingrand, M. P. Georgeff, and A. S. Rao. An architecture for real-time reasoning and system control. *IEEE Expert*, 7(6):34–44, 1992.
17. N. R. Jennings. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial Intelligence*, 75:195–240, 1995.
18. N. R. Jennings. Agent-based computing: Promise and perils. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pages 1429–1436, 1999.
19. J. J. Meyer. Agent languages and their relationship to other programming paradigms. In J. P. Müller, M. P. Singh, and A. S. Rao, editors, *Intelligent Agents V*, pages 309–316. Springer Verlag, Berlin, 1998.
20. P. Noriega and C. Sierra. Towards layered dialogical agents. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III*, pages 173–188, Berlin, 1996. Springer Verlag.

21. S. Parsons and P. Giorgini. An approach to using degrees of belief in BDI agents. In B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, editors, *Information, Uncertainty, Fusion*. Kluwer, Dordrecht, 1999.
22. S. Parsons and N. R. Jennings. Negotiation through argumentation—a preliminary report. In *Proceedings of the International Conference on Multi Agent Systems*, pages 267–274, 1996.
23. 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.
24. A. Rao and M. Georgeff. Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, 1991.
25. 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.
26. A. S. Rao and M. P. Georgeff. Formal Models and Decision Procedures for Multi-Agent Systems. Technical Note 61, Australian Artificial Intelligence Institute, 1995.
27. J. Sabater, C. Sierra, S. Parsons, and N. R. Jennings. Using multi-context agents to engineer executable agents. In *Proceedings of the 6th International Workshop on Agent Theoreies, Archiectures and Languages*, 1999.
28. J. Sabater, C. Sierra, S. Parsons, and N. R. Jennings. Using multi-context agents to engineer executable agents. In N. R. Jennings and Y. Lesperance, editors, *Intelligent Agents IV*, pages 277–294. Springer-Verlag, 2000.
29. J. Sabater, C. Sierra, S. Parsons, and N. R. Jennings. Engineering executable agents using multi-context systems. *Journal of Logic and Computation*, 2002. (to appear).
30. Y. Shoham. Agent-oriented programming. *Artificial Intelligence*, 60:51–92, 1993.
31. S. R. Thomas. The PLACA agent programming language. In M. J. Wooldridge and N. R. Jennings, editors, *Intelligent Agents*, pages 355–370. Springer Verlag, Berlin, 1995.
32. J. Treur. On the use of reflection principles in modelling complex reasoning. *International Journal of Intelligent Systems*, 6:277–294, 1991.
33. L. Vila. *On temporal representation and reasoning in knowledge-based systems*. IIIA Monographies, Barcelona, Spain, 1994.
34. D. Weerasooriya, A. Rao, and K. Rammamohanarao. Design of a concurrent agent-oriented language. In M. J. Wooldridge and N. R. Jennings, editors, *Intelligent Agents*, pages 386–402. Springer Verlag, Berlin, 1995.
35. M. Wooldridge. A knowledge-theoretic semantics for Concurrent MetateM. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III*, pages 357–374. Springer Verlag, Berlin, 1996.
36. M. Wooldridge. Agent-based software engineering. *IEE Proceedings on Software Engineering*, 144:26–37, 1997.
37. M. Wooldridge. *Reasoning about rational agents*. MIT Press, Cambridge, MA, 2000.
38. M. J. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10:115–152, 1995.