Towards an Architecture for Emotional BDI Agents

David Pereira*, Eugénio Oliveira†, Nelma Moreira*, Luís Sarmento†
*DCC – FC & LIACC, University of Porto
Rua do Campo Alegre, 823
4150-180 Porto, Portugal
{dpereira,nam}@ncc.up.pt
†NIAD&R – FEUP & LIACC, University of Porto
Rua Dr. Roberto Frias
4200-465 Porto, Portugal
{eco,las}@fe.up.pt

***Abstract*— In this paper we present the Emotional-BDI architecture, an extension to the BDI architecture supporting Artificial Emotions and including internal representations for agent's Capabilities and Resources. The architecture we present here, is conceptual, defining which components should exist so that Emotional-BDI agents can use Effective Capabilities as well as Effective Resources in order to better cope with highly dynamic environments.**

I. Introduction

The BDI architecture is one of the most well known and studied software agents' architectures. The main reasons are, most notably, the architecture's widely accepted philosophical roots [1], [2], its logical frameworks for modelling and reasoning about BDI agents [3] and a considerable set of software systems which employ the architecture's concepts [4], [5].

More recently, the attention of the Agent Community started to focus on a new (general) kind of agency model: Emotional Agents, which use results about the beneficial aspects of human emotions [6] and apply artificial versions of them in Rational Agents.

In this paper we present the result of a research work aimed at extending the classic BDI architecture with Artificial Emotions: the conceptual Emotional-BDI architecture. These Artificial Emotions are methaphors for the emotions that can be computationally implemented. As a support for these Artificial Emotions, the architecture is equipped with an internal representation of Resources and Capabilities, making the agents "Self-Aware". The main idea is to use Artificial Emotions to help the BDI architecture to incorporate a more accurate model of Practical Reasoning [1] through the interconnection of the mechanisms which are charged of managing the Emotional State of an agent, the Resources and Capabilities, and all the mechanisms which compose the original BDI architecture. This way, the kind of processing done in the architecture we propose is the same as the BDI architecture, except for the fact that there can be applied different kinds of algorithms for performing the same tasks, and these algorithms will be chosen depending on the agents Emotional State, Resources and Capabilities.

This paper is organised as follows. In Section II we will present, from our perspective, the positive and negative aspects of the BDI architecture and draw some conclusions about its real usability in today's problem-solving reality. In Section III we establish a connection between the problems detected and the solutions provided by the functional roles of emotions which deal with Resources and Capabilities management. In Section IV we will start by introducting all the concepts which we consider as prerequisites to the development of a conceptual Emotional-BDI architecture, present this architecture and describe its new components, the interactions between all the architecture's components in terms of information flux and processing, and some topics about implementation. Finally, in Section V we will draw some conclusions about the work to be presented in this paper and also point out the paths for future research.

II. Motivation

The BDI architecture has shown to be a very successfull one, as it can be seen from the number of software systems which implemented it [4], [5], [7]. However, we believe that the architecture can still be more effective so that it can face the new generation of problems characterised by high levels of unpredictability, complexity and dynamics. Some opinions about the necessity of this enhancement process were already put forward by [3] and, therefore, we followed those clues to conduct our own research.

A. Pros

First, let us review the main reasons for the BDI architecture's success in the Agents Community. These can be divided in three different classes [3]:

Strong philosophical roots: The BDI Architecture was created by Bratman et al. [1] with the goal of being able to establish a good balance between reactive and deliberative behaviour, while still being a pure deliberative architecture. For that, and in addition to the philosophical concept of Intentional System [2] that underlies the own notion of agent, the architecture uses the concept of Resource Bounded Practical Reasoning which is a variant of the classical reasoning directed towards the execution of actions, but where Intentions have a central role in driving means-ends reasoning, constraining deliberation and influence the Belief base of an agent [1], [8]. Both philosophical concepts are widely accepted and also provide a natural way of describing the behaviour which agents should exhibit only through the use of the following mental states: Beliefs, Desires and Intentions. Therefore, the development of BDI agents is feasible either by computer programming experts or experts of the domain where the agents will act.

Elegant logical formalisms: Another attractive aspect of the BDI architecture is the set of logical frameworks exclusively developed to reason about BDI agents, and which also provide an important guidance towards the correct development of software programs which make use of the architecture. The most well know are the Rao and Georgeff's $\mathbf{BDI}_{\mathsf{CTL}^*}$ framework and the $\mathbf{KARO}$ framework of Woek et al. For a more detailed overview about these two logical frameworks refer to [9].

Software implementations: The BDI architecture has been applied to software systems on the realms of academic and real-world software systems. In the realm of academic research, programming languages were developed which embody the BDI model and thus diminish the amount of *ad hoc* coding, such as the PRS, dMARS, AgentSpeak and 3APL [10]. Also in the realm of academic research there exists a considerable set of agent programming frameworks which provide a set of tools for an easier and consistent development of BDI agents. Well known and used frameworks are JACK [10], Jadex BDI Agent System [11] and Jason [12]. On the realm of real-word applications, the BDI architecture was applied with a great amount of success. The most important application of this architecture is Georgeff's project for diagnose faults in the reaction control system of the Shuttle Discovery [5], although there are other examples [4], [7] of the validity of this architecture in software implementations.

B. Cons

Lets now analyse the main problems associated with the usage of the BDI architecture for the development of agent-based systems.

Lack of information about resource bounds: The BDI architecture uses only specific roles of Intentions [8] to control the problem of acting under resource bounds. In our point of view this does not seems to be enough, since today's problems are characterised by a continuously growing of complexity and unpredictability, under severe resource bounds. We believe that the BDI model lacks an explicit internal representation of the means which an agent can count on in order to decide which is the best way of acting on its environment without unnecessarily compromising future actions and also its overall performance.

The problem of agents reconsideration: This problem is the consequence of the relation that exists between an environment's change rate and the frequency and amount of reconsideration (deliberation) which BDI agents which inhabit this environment do. Since the frequency of reconsideration carried out by BDI agents must be defined in advance, these may miss important changes that occur in the environment or may carry out unfruitful reconsideration, in the case of not occurring significant changes on the environment. Therefore, we argue that both the deliberation processes and the instruments used in them should be dynamic and as much adaptive to environment's changes as possible.

Lack of other human-like mental states: The description of entities and the prediction of their behaviour under the rules of Dennet's Intentional Stance [2] is not limited to the usage of Beliefs, Desires and Intentions. Despite the fact that these mental states are flexible enough to model BDI agents for acting in a wide range of scenarios, there are some cases where the usage of other mental states would be appropriate [13]. The usage of other mental states would provide to computer scientists and domain specific experts, respectively, new computational structures to implement software agents and new concepts for developing more refined models of BDI agents.

C. Comparing the pros and cons

From the above results we have concluded that the BDI architecture is still being a valuable concept to have in account when the development of software systems which requires Rational Agents is the case. However, this architecture suffers from problems which are far from being ignored and urge to be solved. In our opinion the first attempt to be made should fall upon the additional usage of accurate and specialised components to deal with the explicit resources which agents have, both for their usage in reconsideration processes and in all the functions which are part of the architecture's processing cycle [8]. Wooldridge et al. also reached a similar conclusion some years ago [3], when they have proposed the development of specialised heuristics for dealing with the problems encountered in the BDI architecture. The components we are proposing are surely elements of a set composed by such heuristics.

III. Artificial Emotions, Agent's means and the BDI architecture

The usage of Artificial Emotions in the BDI architecture is not new [14]. However, the idea of using them as mechanisms for controlling the means of the agents for acting upon their environment is new, at least for what we are aware.

From the set of the three fragilities we have found in the BDI architecture, only one is not fully dependent on the addition of new concepts to the architecture: the *lack of other human-like mental states* can be solved by using the same approach used for deciding that Beliefs, Desires and Intentions would form the base of the BDI architecture. In other words, we can use the Intentional Stance [2] and add commonsense definitions of new mental sates, such as emotions, and make them influence the BDI architecture through the commonsense understanding of the way they affect positively the reasoning performed by humans. For instance, we can define a new concept such as Fear wich is something like an informational data structure which reports situations which an agent should avoid.

The problems of the *lack of information about resource bounds* and the *problem of agents reconsideration* requires a more refined approach, in terms of the usage of Artificial Emotions, mapping the way they are activated and which kind of plans of action they trigger. Consider, for instance, the problem of the amount and frequency of reconsideration done by an agent when it faces a threat. It seems acceptable that, instead of reactivelly going away, it should apply the maximum of its resources and allocate them to the best solution-search algorithms it possesses, thus finding the best solution for escaping the threat. In fact, this is the role of driving attention and self-awareness updating that emotions play in humans. If an agent acts upon rules similar to the previous one, the reconsideration process may become dynamic (both on the amount of the means used and on the frequency of its execution) and adaptive, thus making the agent respond better to the changes ocurring in the environment.

For the application of such clues that emotions present in humans give to us, on the perspective of a correctly usage of the available means, we have also to have a clear representation of these kind of means. Therefore, this leads us to the problem of the lack of information about resource bounds. This kind of information influences the Emotional State of the agent, but also is influenced by it, since the emotional state affects the way perception is achieved. The architecture we present tries to link all these concepts by giving an abstract idea of how they should be defined and how they should interact.Therefore, we will present only a conceptual architecture, leaving more low-level details to future implementations which will be conducted for analysing with more detail the validity of our approach, relatively to other architectures.

IV. The Emotional-BDI architecture

We will now describe the proposed extension of the BDI architecture for supporting Artificial Emotions. We will start by introducing the basic concepts for such extension, followed by the conceptual architecture itself and also an abstract interpreter which maps the architecture's processing cycle.

A. Effective Resources and Effective Capabilities

The need for an explicit internal representation of the means that an agent has to execute upon its environment has been stress as the basis for the extension we are proposing. We argue that the concepts of Resource and Capability should form the basis of such internal representation.

Capabilities: These are abstract plans of action which the agent has available to act upon its environment. In terms of agent conception, a set of Capabilities can be seen as a dynamic Plan Library, where some of its plans are marked as being impossible to be considered either temporarily or permanently. Padgham and Lambrix introduced the idea of Capabilities in the BDI architecture [15]. Although they consider as Capability both the *ability* and *opportunity* and therefore an agent does not depend on any other resource to engage their execution, they enforce the idea that a Capability can be considered as a Plan which may not always be available.

Resources: These are means that turn Capabilities into plans of action which can be performed by the agent in its environment. Resources can be either physical (CPU time, disk space, available memory, etc.) or virtual (energy sources on a virtual world, other agents, etc), or both. They are also the source of the dynamics associated with the Capabilities, since the availability or unavailability of the Capabilities depends on the existing amount of Resources.

Agents are not omniscient entities, thus only have information about a limited part of their environment and about itself. In particular this also applies to their Resources and Capabilities. Thus, what an agent can count on is its Effective Resources, which are a subset of all the available Resources which the agent is aware of. Capabilities, when instantiated to the required Effective Resources become Effective Capabilities.

The process of both Resources and Capabilities become effective is done by the following function which we decided to call *Effective Capabilities and Effective Resources revision function*, or $EC\text{-}ER\text{-}rf$. The behaviour of this function is as follows:

- $EC\text{-}ER\text{-}rf$:$\mathbf{Cap} \times \mathbf{Res} \times \mathbf{Percept} \times \langle \mathbf{Cap} \times \mathbf{Res} \rangle \rightarrow \langle \mathbf{Cap} \times \mathbf{Res} \rangle$
  This function revises and updates the Resources and Capabilities which in fact are available for an agent to use. The information that carries data needed for this process comes from the data percepted by the agent and also from the various functions of the BDI architecture, which report the amount of Resources consumed, and which of the Capabilities were used. The result of this evaluation is forwarded to the mechanism responsible for the Emotional State update, so an agent become adapted to the new reality of the means it has.

A schema of the entities which forms this component is presented in Figure 1.

B. Sensing and Perception

In order to interpret the meaning of the stimuli that occurs in an environment, and that is captured by the agents sensorial machinery, we need to equip these agents with mechanisms which have rules that binds these stimuli to concepts which "make sense" in these agents reasoning procedures.

We propose a *Sensing and Perception Module*, a mechanism capable of obtaining the desired information from stimuli provided by an environment. This module is composed of two sub-mechanisms responsible for performing different, but complementary tasks: a *sensing filter* which deals directly with

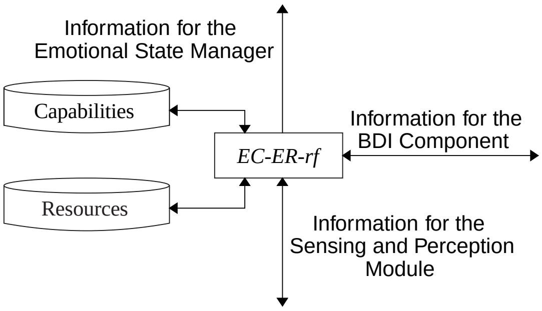

Fig. 1. The Effective Capabilities Manager module.

the stimuli captured by sensors of the agent; and a *perception filter*, which attributes meaning to previous filtered data. In more detail we have:

sensing filter: this filter is responsible for extracting information directly from the stimuli captured by an agent's sensorial machinery. For that, it is connected to a repository of information extraction rules which can be directly applied to the information provided by the sensors. The rules which are applied are dynamically chosen, depending on the Effective Capabilities and on the Emotional State of the agent. Therefore, the kind of information extracted for further processing will emulate the role of driving attention that real emotions play in human reasoning.

perception filter: this filter is responsible for giving meaning to data previously processed by the *sensing filter*. In order to attribute a semantics for chunks of data, this mechanism uses a repository of *semantic association rules*. The result of the application of this filter is the production of concepts which can be forwarded for both *belief revision processes* of the agent, and also for *Effective Resources and Effective Capabilities revision*. Also the kind of rules that are available for being applied depends on the Effective Capabilities and Emotional State of the agent.

An illustrative example of how meaning is attributed to stimuli is presented in Figure 2. The full *Sensing and Perception Module* is depicted in Figure 3.

C. The Emotional State Manager

The Emotional State Manager is the component responsible for controlling the Resources and the Capabilities used in the information processing phases of the architecture. It should be composed by an internal structure which exhibits an adaptive behaviour, depending on the changes that occur in the environment where agents are standing. Thus, in order to avoid falling into a problem near to the problem of reconsideration detected in the BDI architecture, we do not present any fixed structure which this Emotional State Manager should incorporate. Instead, we only propose the following conduits that we consider to be fundamental for this kind of component:

- it should base itself on a well defined set of *Artificial Emotions* which relates efficiently the kind of tasks the agent has to perform, and the way changes on the environment affect the internal state of the agent. For instance, if the primary goal of an agent is survival, then Fear should be present in the Emotional Set so the agent can be alerted when its survival changes diminish considerably.
- there should exist various information extraction functions for each of the *Artificial Emotions*. This dues to the fact that there are various sources of stimuli which change the same emotion in different ways. For instance, for a Fireman which is fighting a fire, the Fear should be elicited both by fire proximity and changes on the wind, which may change the fire's direction.
- there should exist a decay rate for each of the *Artificial Emotions*, but always depending on the state of the emotional eliciting source. In the example above, if the Fireman continuously notices fire near him, he will continue in a Fear emotional state. However, if it escapes the fire, the Fear rate will diminish with time.

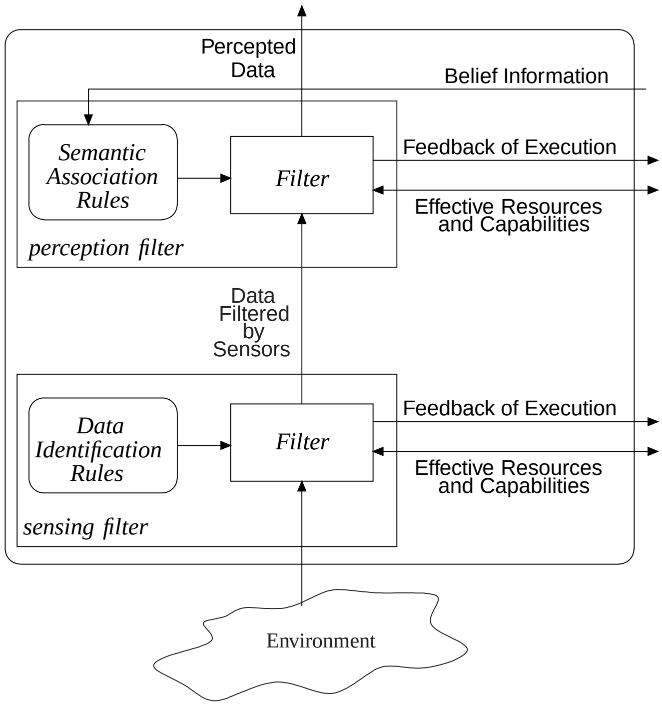

Fig. 3. The Sense and Perception manager.

An interesting model for Emotional State managing can be seen on the work of Oliveira and Sarmento [16].

D. The conceptual architecture

The Emotional-BDI architecture we are proposing is an extended version of the classic BDI architecture with the addition of three new components – the *Sensing and Perception Manager*, the *Effective Capabilities and Effective Resources revision function* and the *Emotional State Manager* – and the

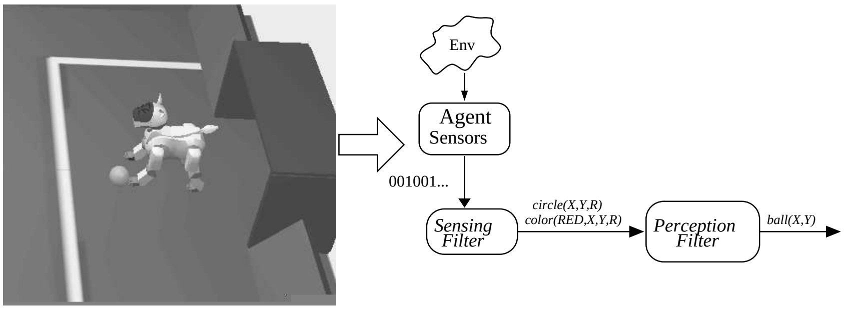

Fig. 2. Identification of a stimuli as a ball of a robotic soccer match.

definition of interactions between these new components and between these new components and the old ones. We present the overall conceptual Emotional-BDI architecture shown in Figure 4.

The interactions between the new components have already been described in Sections IV-A, IV-B and IV-C. We need now to describe the interaction between these new components and the ones which are present in the original BDI architecture.

The BDI functions of our architecture have the same name as the original ones plus an apostrophe concatenated at the end of their name. We denote the Beliefs by Bel, the Desires by Des, the Intentions by Int, the Effective Resources by $\mathbf{ERes}$, the Effective Capabilities by $\mathbf{ECap}$, the information percepted by $\mathbf{Percept}$, and finally the Emotional State by $\mathbf{EState}$. These new components are the following:

brf'**:** $\mathsf{Bel} \times \mathbf{Percept} \times \mathbf{ECap} \times \mathbf{ERes} \times \mathbf{EState} \rightarrow \mathsf{Bel}$
: This function updates the Beliefs of the agent based on existing Beliefs and on the new information percepted by the *Sensing and Perception Module*. The computational resources spent and the revision algorithms used are defined by an evaluation made on the base of the Effective Resources, Effective Capabilities and the Emotional State of the agent.

$gen\text{-}options'$**:** $\mathsf{Bel} \times \mathsf{Int} \times \mathbf{ECap} \times \mathbf{ERes} \times \mathbf{EState} \rightarrow \mathsf{Des}$
: This function is responsible by means-ends reasoning, the process of recursively elaborating hierarchical plans which defines progressively more specific Intentions, until these become satisfiable by the execution of actions. This function is also controlled by both the Effective Resources/Capabilities and the Emotional State which enables the use of distinct algorithms, making it adaptive to the changes of the environment.

$filter'$**:** $\mathsf{Bel} \times \mathsf{Des} \times \mathsf{Int} \times \mathbf{ECap} \times \mathbf{ERes} \times \mathbf{EState} \rightarrow \mathsf{Int}$
: This function updates the Intentions based on already existing Intentions, Beliefs and Desires. Through it, Intentions should be dropped if they were already achieved or if they will never be achievable, and should also retain the ones which are considered to be fruitfull in the future. Once again, distinct kinds of algorithms may be used, depending on the Emotional State of the agent.

$execute'$**:** $\mathsf{Int} \times \mathbf{ECap} \times \mathbf{ERes} \times \mathbf{EState} \rightarrow \mathsf{Action}$
: This function selects an Intention to be fulfilled by the direct excution of an action. The action and the Intention selection can be a dynamic process, depending on the available Effective Resources, Effective Capabilities and the Emotional State.

Besides having the functionality that was described, each of the above functions also informs about the subset of Effective Resources and Effective Capabilities used so that the EC-ER-rf function can update correctly what is left to be used. These functions also informs the Emotional State Manager about their overall performace, in order to update the Emotional State of the agent. These analysis can contain, for instance, the number of solutions obtained, the time took to obtain them, which algorithms were used, if errors were detected, etc.

Based on the schematic view of the conceptual Emotional-BDI architecture and on the properties of each of the above functions, we present an abstract interpreter for our architecture.

```
Emotional-BDI-interpreter
initialize-emotional-state-manager(ES);
initialize-sensing-perception-module(ES);
initialize-EC-ER-rv(C,R);
initialize-bdi-state();
repeat
  options := option-generator(sensor-input,B,D,I,EC,ER,ES);
  selected-options := deliberate(options,B,D,I,EC,ER,ES);
  execute(I,EC,ER,ES);
  get-new-external-events(sensing-perception-filter,EC,ER,ES);
  drop-successful-attitudes(B,D,I,EC,ER,E);
  drop-impossible-attitudes(B,D,I,EC,ER,ES);
end repeat
```

Fig. 5. The Emotional-BDI abstract interpreter.

The abstract interpreter (Figure 5) is very similar to the original one [8], except that the various components are initialised separately and all the functions include as input the Effective Capabilities, the Effective Resources and the Emotional State of the agent. Moreover, this interpreter enforces the idea that our architecture works as one component only, as stated in [6].

V. Conclusions and Future Work

In this paper we have presented a conceptual Emotional-BDI architecture, an extension to the original BDI architecture with the addition of an internal representation of the means

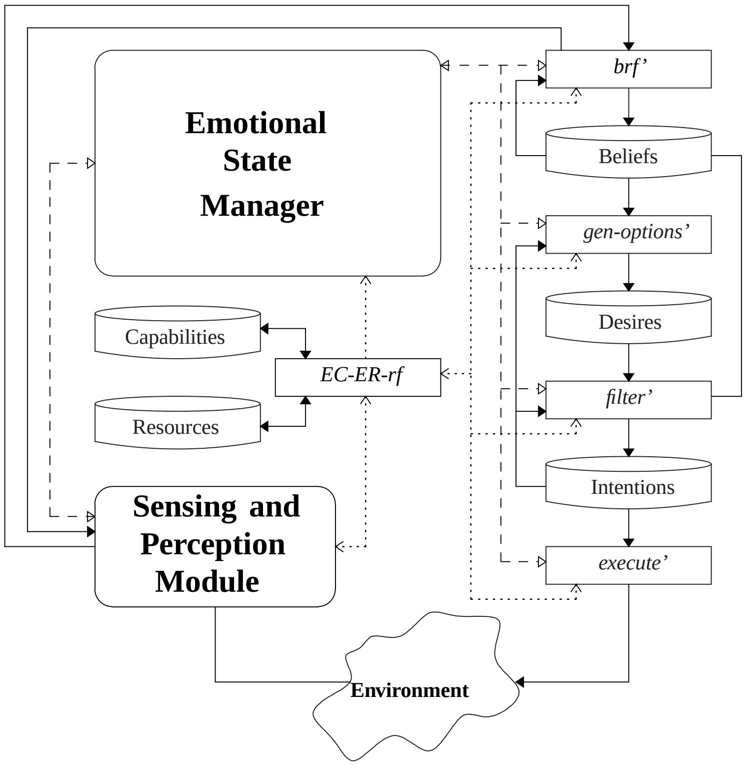

Fig. 4. The conceptual Emotional-BDI architecture.

an agent can count on, and an Emotional State capable of controlling the usage of these means in order for the agent to perform adaptively in its environment. We believe that both the components fill the gaps which we argued that exist in the BDI architecture.

Our future work will be on the implementation of Emotional-BDI agents in dynamic and complex environments like PyroSim [16], but also in static environments like a simple maze with energy sources and obstacles, possible of being developed, for instance, in freeBots[1].

Parallel to the implementation issues we have just referred, we will try to extend the existing logical frameworks to support the concepts present in the Emotional-BDI architecture (and the implementation issues itselves) with the intent of avoiding unpleasant *ad hoc* coding techniques and to have the possibility to verify properties of our architecture.

Acknowledgment

This work was partially founded by Fundação para a Ciência e Tecnologia (FCT) and program POSI.

References

[1] M. E. Bratman, D. Israel, and M. E. Pollack, "Plans and resource-bounded practical reasoning." *Computational Intelligence*, vol. 4, pp. 349–355, 1988.

[1] freeBots is a robot simulation framework that is being developed at DCC-FC & LIACC, by Luís Damas.

[2] D. C. Dennett, *The Intentional Stance*. MITP, 1987.
[3] M. P. Georgeff, B. Pell, M. E. Pollack, M. Tambe, and M. Wooldridge, "The belief-desire-intention model of agency." in *ATAL*, ser. Lecture Notes in Computer Science, J. P. Müller, M. P. Singh, and A. S. Rao, Eds., vol. 1555. Springer, 1998, pp. 1–10.
[4] M. P. Georgeff and A. S. Rao, "A profile of the australian artificial intelligence institute." *IEEE Expert*, vol. 11, no. 6, pp. 89–92, 1996.
[5] M. Georgeff, "The mind of the machine," http://www.nasa.gov.
[6] A. R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, New-York, Ed. Grosset/Putnam, 1994.
[7] A. S. Rao and M. P. Georgeff, "BDI-agents: from theory to practice," in *Proceedings of the First Intl. Conference on Multiagent Systems*, San Francisco, 1995. [Online]. Available: citeseer.ist.psu.edu/rao95bdi.html
[8] G. Weiss, Ed., *Multiagent systems : a modern approach to distributed artificial intelligence*. Cambridge, Mass.: MIT Press, 1999.
[9] W. van der Hoek and M. Wooldridge, "Towards a logic of rational agency." *Logic Journal of the IGPL*, vol. 11, no. 2, pp. 135–159, 2003.
[10] D. D. Davide Ancona and V. Mascardi, "A survey on languages for programming bdi-style agents."
[11] A. P. Lars Braubach, "Jadex bdi agent system," http://vsis-www.informatik.uni-hamburg.de/projects/jadex/, May 2005.
[12] R. H. Bordini and J. F. Hübner, "Jason – a java-based agentspeak interpreter used with saci for multi-agent distribution over the net," http://jason.sourceforge.net/, March 2005.
[13] E. Norling, "Folk psychology for human modelling: Extending the bdi paradigm." in *AAMAS*. IEEE Computer Society, 2004, pp. 202–209.
[14] L. Padgham and G. Taylor, "A system for modelling agents having emotion and personality." in *PRICAI Workshop on Intelligent Agent Systems*, ser. Lecture Notes in Computer Science, L. Cavedon, A. S. Rao, and W. Wobcke, Eds., vol. 1209. Springer, 1996, pp. 59–71.
[15] L. Padgham and P. Lambrix, "Formalisations of capabilities for bdi-agents," *Autonomous Agents and Multi-Agent Systems*, no. 10, pp. 249–271, 2005.

[16] E. Oliveira and L. Sarmento, “Emotional advantage for adaptability and autonomy.” in *AAMAS*. ACM, 2003, pp. 305–312.