Modelling Emotional BDI Agents

David Pereira[1], Eugénio Oliveira[2], and Nelma Moreira[1]

[1] DCC – FCUP & LIACC, Rua do Campo Alegre, 823, 4150-180 Porto, Portugal
{dpereira,nam}@ncc.up.pt

[2] NIAD&R – FEUP & LIACC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
eco@fe.up.pt

Abstract. Emotional-BDI agents are agents whose behaviour is guided not only by beliefs, desires and intentions, but also by the role of emotions in reasoning and decision-making. In this paper we introduce the logic $\mathcal{EBDI}$ for specifying Emotional-BDI agents in general and a special kind of Emotional-BDI agent under the effect of fear. The focus of this work is in the expressiveness of $\mathcal{EBDI}$ and on using it to establish some properties which agents under the effect of an emotion should exhibit.

1 Introduction

Emotional-BDI agency describes computational agents whose behaviour is guided by the interactions existing between beliefs, desires and intentions (along the lines of the classical BDI architecture [1]), but where these interactions are influenced by a third-party emotional component [2]. This component produces data which will bound the BDI interaction by imposing some of the large set of positive aspects that emotions play in reasoning and decision-making [3].

The conceptual architecture which defines the Emotional-BDI model of agency was recently introduced in [2] and is mainly based on the work of emotional agents Oliveira & Sarmento's emotional agent architecture [4], although adapted to fit in the original BDI architecture [1, 5, 6].

In this paper we introduce $\mathcal{EBDI}$, a multi-modal logic for specifying Emotional-BDI agents. We define the various axioms which properly characterise each of the modal operators of $\mathcal{EBDI}$ and after we give the specification of the basic Emotional-BDI agent and a specification of a *fearful* Emotional-BDI agent.

This paper is organised as follows: in Section 2 we provide the motivation for the work we are presenting; in Section 3 we introduce the logic $\mathcal{EBDI}$ and define its syntax, semantics and axioms for the modal operators; in Section 4 we present the specification of a basic Emotional-BDI agent and a *fearful* Emotional-BDI agent. Finally, in Section 5 we refer similar work of other authors and in Section 6 we draw some conclusions and point the path to current and future work.

2 Motivation

The main motivation for the current work was to provide a formal system in which the concepts of the Emotional-BDI model of agency could be logically expressed. Using these concepts, we can build distinct specifications of Emotional-BDI agents which describe the behaviours which are expected from the agents under the influence of emotions.

The existing formal systems, namely $\mathsf{BDI}_{\mathbf{CTL}}$ [6] and the **KARO** [7, 8] framework, if used independently, are not suited for our goals. However, both have properties which we need to combine in order to properly model Emotional-BDI agents.

Plus, we integrate some important concepts of Oliveira & Sarmento's emotional agent architecture [4], which were mapped into abstract concepts for fitting the structure of $\mathcal{EBDI}$'s syntax.

3 The Logic $\mathcal{EBDI}$

We will now introduce the logic $\mathcal{EBDI}$. We first give a resumed informal description of the purpose of each of its components and afterwards we provide its syntax and semantics.

3.1 Informal semantics

The logical structure which supports $\mathcal{EBDI}$ is a branching-time temporal structure introduced by Schild [9], which is a simplified approach to Rao & Goergeff's $\mathsf{BDI}_{\mathbf{CTL}}$ [10] semantics. The temporal relation is established over a set of elements, called *situations*. Situations are pairs $\langle world, state \rangle$, where the *world* refers to a scenario the agent considers as valid, and the *state* refers to a particular point in that scenario.

In $\mathcal{EBDI}$ we consider explicit complex actions, as in the **KARO** framework. Actions can be either atomic or regular: the first are actions which cannot be sub-divided into a combination of smaller ones, while regular actions are constructions of atomic actions through some set of regular rules. Actions are a labelling of the branching time structure underlying $\mathcal{EBDI}$.

In order to properly execute any action, we need the notion of capability (abstract plan) already studied in [11, 7] and also the explicitly notion of *resource*. We use these to specify under which conditions the agent is able to effectively execute any action.

Finally, we introduce the concepts of fear and fundamental desire. The first refer to *fearing something* or being *fearful that*, and brings concepts into objects of fear in $\mathcal{EBDI}$. To properly establish the notion of fear, we require to have special information in which are described the vital desires of an agent, like, for instance, to be alive. The notion of fundamental desire plays such a role. Although it is a desire, a fundamental desire has special properties which guarantee the existence of the agent in an environment.

3.2 Syntax

We now define the language of $\mathcal{EBDI}$ which extends Rao & Georgeff's $\mathsf{BDI}_{\mathbf{CTL}}$ [10] for containing explicit actions, capabilities, resources and modal operators representing fear and fundamental desires. This language distinguishes between *state-formulas* (which are evaluated in a given situation) and *path-formulas* (which are evaluated along a given path).

Definition 1. *Given an infinite numerable set* $P = \{p, q, p_1, \ldots\}$ *of propositional variables and an infinite numerable set of atomic actions* $A_{\mathsf{At}} = \{a, b, a_i, \ldots\}$*, the set of* $\mathcal{EBDI}$ well-formed formulas *defined by the following BNF-grammar:*

- *State-formulas (SF):*
  $\varphi_s ::= p \mid \neg\varphi_s \mid \varphi_s \wedge \varphi_s \mid$
  $[\alpha]\varphi_s \mid \langle\alpha\rangle\varphi_s \mid \mathbf{E}\varphi_p \mid \mathbf{A}\varphi_p$
  $\mathsf{BEL}(\varphi_s) \mid \mathsf{DES}(\varphi_s) \mid \mathsf{INT}(\varphi_s) \mid \mathsf{FEAR}(\varphi_s) \mid \mathsf{FDES}(\varphi_s) \mid$
  $\mathsf{CAP}(\alpha) \mid \mathsf{RES}(\alpha)$
- *Path-formulas (PF):*
  $\varphi_p ::= \mathbf{X}(\varphi_s) \mid \varphi_s \mathbf{U} \varphi_s$
- *Regular-actions (*A_{Ra}*):*
  $\alpha ::= id \mid a_i \mid \alpha;\alpha \mid \alpha + \alpha \mid \alpha^*$

In addition, we introduce the following abbreviations: $\top$, $\bot$, $\varphi \vee \psi$ and $\varphi \rightarrow \psi$ are abbreviations of $\neg(p \wedge \neg p)$ (with p being a fixed element of P), $\neg\top$, $\neg(\neg\varphi\wedge\neg\psi)$ and $\neg\varphi\vee\psi$, respectively; $\mathbf{AF}\varphi$, $\mathbf{EF}\varphi$, $\mathbf{AG}\varphi$ and $\mathbf{EG}\varphi$ are abbreviations of $\mathbf{A}(\top\mathbf{U}\varphi)$, $\mathbf{E}(\top\mathbf{U}\varphi)$, $\neg\mathbf{EF}\neg\varphi$ and $\neg\mathbf{AF}\neg\varphi$, respectively. Iterated actions are inductively defined by $\alpha^0 = id$ and $\alpha^{n+1} = \alpha;\alpha^n$.

3.3 Semantics

In this section we introduce the semantics of $\mathcal{EBDI}$. We start by defining the notion of situation.

Definition 2. *Given an non-empty set* $W = \{w_0, w_1, w_2, \ldots\}$ *of worlds (also known as agent's perspectives or scenarios), and a non-empty set* $S = \{t_0, t_1, t_2, \ldots\}$ *of temporal-states (also known as time points), a* situation *is a pair* $\sigma = \langle w_i, s_j\rangle$*, with* $i \geq 0$ *and* $j \geq 0$*. The set of situations is denoted as* Σ*, which verifies* $\Sigma \neq \emptyset$ *and* $\Sigma \subseteq W \times S$*.*

Situations define particular temporal states, in scenarios that the agent has information about. For instance, in a situation $\langle desire, t\rangle$ the desire of winning the lottery may be considered as true, although in the same temporal state, lets say $\langle belief, t\rangle$, the agent may not believe in it. However, at some temporal state t' both may be considered true by the agent.

Given a set of situations Σ we can map the evolution of time and action execution by defining two relations: one is a *branching time* relation

Definition 3. *Given a non-empty set of situations* Σ *we define the* relation $\mathcal{R}_T$ *as follows:*

1. *It is serial, i.e.,* $\forall \sigma \in \Sigma$, $\exists \sigma' \in \Sigma$ *such that* $(\sigma, \sigma') \in \mathcal{R}_T$*;*
2. *If* $(\langle w_i, s_j \rangle, \langle w_k, s_l \rangle) \in \mathcal{R}_T$ *then* $w_i = w_k$*.*

and the other is a *action execution* relation that associates to each element of $\mathcal{R}_T$ an atomic action.

Definition 4. *Given a set of atomic-actions* A_{At} *and a branching time relation* $\mathcal{R}_T$*, for* $a_i \in A_{\mathsf{At}}$ *we define* $\mathcal{R}_{a_i}$ *such that:*

1. *If* $\mathcal{R}_{a_i} \in \mathcal{R}_T$*;*
2. *If* $(\sigma, \sigma') \in \mathcal{R}_{a_i}$*, then it is false that exists* $a_j \in A_{\mathsf{At}}$ *such that* $i \neq j$ *and* $(\sigma, \sigma') \in \mathcal{R}_{a_j}$*;*

The previous relation can be extended to regular actions, as follows.

Definition 5. *Given a regular action* α *and a set of situations* Σ*, we inductively define the* regular action accessibility relation *by:*

$$
\begin{array}{ll}
R^A & : \; A_{\mathsf{Ra}} \to (\Sigma \times \Sigma) \\
R^A(a_i) & = \{(\sigma, \sigma') \mid (\sigma, \sigma') \in \mathcal{R}_{a_i}\} \\
R^A(id) & = \{(\sigma, \sigma') \mid \sigma = \sigma'\} \\
R^A(\alpha;\beta) & = \{(\sigma, \sigma') \mid \exists \sigma'' \in \Sigma((\sigma, \sigma'') \in R^A(\alpha) \wedge (\sigma'', \sigma') \in R^A(\beta))\} \\
R^A(\alpha+\beta) & = \{(\sigma, \sigma') \mid (\sigma, \sigma') \in R^A(\alpha) \;\; \mathit{or} \; (\sigma, \sigma') \in R^A(\beta)\} \\
R^A(\alpha^0) & = \{(\sigma, \sigma') \mid (\sigma, \sigma') \in R^A(id)\} \\
R^A(\alpha^{(n+1)}) & = \{(\sigma, \sigma') \mid (\sigma, \sigma') \in R^A(\alpha;\alpha^n)\} \\
R^A(\alpha^*) & = \{(\sigma, \sigma') \mid \exists n \in \mathbb{N}((\sigma, \sigma') \in R^A(\alpha^n))\}
\end{array}
$$

The main interest behind using both approaches to is mainly guided by the properties which emotions exhibit. The emotions can be triggered either by an action which will lead to some wanted/unwanted situation or triggered by believing that such situations may or will inevitably be true in the future.

The distinction, in the syntax, between path formulas and state formulas must reflect also in the semantics. In $\mathcal{EBDI}$, as in $\mathsf{BDI}_{\mathbf{CTL}}$, the former are analysed along a path and the second in a particular situation. In $\mathcal{EBDI}$, a path is defined as follows:

Definition 6. *Let* Σ *be a set o situations and* $\mathcal{R}_T$ *a branching time relation defined on* Σ*. A* path *is a subset* $\pi_\sigma = (\sigma_0, \sigma_1, \sigma_2, \ldots)$ *such that* $\sigma = \sigma_0$ *and* $\forall i \geq 0$*,* $(\sigma_i, \sigma_{i+1}) \in \mathcal{R}_T$*. Th* k^{th} *element of a path* π_σ *is denoted as* $\pi_\sigma[k]$*.*

We already saw that we can analyse the several perspectives the agent may be aware of at the same state. For that we have to vary the scenario component of any situation $\langle \mathit{scenario}, \mathit{state} \rangle$. The relations which establish this relationship are the ones which are going to be used for modelling the mental states of the agent and respect the following condition: let $\mathcal{R}$ be a relation, if $(\langle w_i, s_j \rangle, \langle w_k, s_l \rangle) \in \mathcal{R}$ then $s_j = s_l$.

Finally, we also have to provide a semantic interpretation for capabilities and resources. We mainly follow the ideas of modelling capabilities in the **KARO** framework, which is by considering local functions in each situation which establish which atomic actions the agent has capabilities/resources to execute properly. The capabilities/resources for regular actions are interpreted by relating these local functions to regular action accessibility relations, in the following way.

Definition 7. *Given a regular action α, a set of situations Σ and a function $\mathtt{v}_f(a_i)$ which establishes a subset of Σ where the agent has capabilities/resources to execute atomic actions a_i, resources and capabilities are interpreted by similar functions, just varying in the local functions $\mathtt{v}_f(a_i)$. Therefore, we inductively define them in one function f, with $f \in \{c, r\}$, such that:*

$$
\begin{array}{ll}
f^A & : \ A_{\mathsf{Ra}} \rightarrow \wp(\Sigma) \\
f^A(a_i) & = \mathtt{v}_f(a_i) \\
f^A(id) & = \Sigma \\
f^A(\alpha;\beta) & = \{\sigma \mid \sigma \in f^A(\alpha) \wedge \exists \sigma' \in \Sigma((\sigma, \sigma') \in R^A(\alpha) \wedge \sigma' \in f^A(\beta))\} \\
f^A(\alpha{+}\beta) & = \{\sigma \mid \sigma \in f^A(\alpha) \vee \sigma \in f^A(\beta)\} \\
f^A(\alpha^0) & = \{\sigma \mid \sigma \in f^A(id)\} \\
f^A(\alpha^{(n+1)}) & = \{\sigma \mid \sigma \in f^A(\alpha;\alpha^n)) \\
f^A(\alpha^*) & = \{\sigma \mid \exists n \in \mathbb{N}(\sigma \in f^A(\alpha^n))\}
\end{array}
$$

The interpretation of $\mathcal{EBDI}$-formulae is done over Kripke-models, as defined below.

Definition 8. *Given a set of worlds W, a set of temporal states S, a set of propositional variables P, a set of atomic actions A_{At} and a set of modal operators $Op = \{\mathsf{BEL}, \mathsf{DES}, \mathsf{INT}, \mathsf{FDES}, \mathsf{FEAR}\}$, we define an $\mathcal{EBDI}$-*model *as a tuple*

$$M = \langle \Sigma, \mathcal{R}_T, \{\mathcal{R}_a : a \in A_{\mathsf{At}}\}, R^A, \{\mathcal{R}^{\mathsf{O}} : \mathsf{O} \in Op\}, c^A, r^A, \mathtt{v}_p, \mathtt{v}_c, \mathtt{v}_r \rangle$$

where

- *Σ is the set of situations;*
- *$\mathcal{R}_T$ is a branching time relation on Σ;*
- *each $\mathcal{R}_{a_i}$ is a atomic action accessibility relation on Σ;*
- *R^A is a accessibility relation for regular actions;*
- *$\mathcal{R}^{\mathsf{O}}$ are accessibility relations for the corresponding modal operators;*
- *$\mathtt{v}_p, \mathtt{v}_c$ and $\mathtt{v}_r$ are functions which define in which states the propositions hold, the capabilities for atomic actions hold and the resources for atomic actions hold, respectively.*

The satisfiability of a well-formed formula in $\mathcal{EBDI}$ is given by the following definition.

Definition 9. *Let M be an $\mathcal{EBDI}$-model. The* satisfiability *of a $\mathcal{EBDI}$-formula with respect to M and a situation $\sigma \in \Sigma$ is inductively defined as follows, considering $\mathsf{O} \in Op$:*

- *satisfaction for state-formulas:*
  - *(sf1)* $M, \sigma \models p$ *iff* $p \in \mathtt{v}_p(\alpha)$
  - *(sf2)* $M, \sigma \models \neg\varphi$ *iff* $M, \sigma \not\models \varphi$
  - *(sf3)* $M, \sigma \models \varphi \wedge \psi$ *iff* $M, \sigma \models \varphi$ *and* $M, \sigma \models \psi$
  - *(sf4)* $M, \sigma \models \mathbf{E}\psi$ *iff* $\exists \pi_{\theta_\sigma}$ *such that* $M, \pi_{\theta_\sigma} \models \varphi$
  - *(sf5)* $M, \sigma \models \mathbf{A}\psi$ *iff* $\forall \pi_{\theta_\sigma}$, $M, \pi_{\theta_\sigma} \models \varphi$
  - *(sf6)* $M, \sigma \models \langle\alpha\rangle\varphi$ *iff* $\exists\, (\sigma, \sigma') \in R^A(\alpha)$ *such that* $M, \sigma' \models \varphi$
  - *(sf7)* $M, \sigma \models [\alpha]\varphi$ *iff* $\forall\, (\sigma, \sigma') \in R^A(\alpha)$, $M, \sigma' \models \varphi$
  - *(sf8)* $M, \sigma \models \mathsf{O}(\varphi)$ *iff* $\forall\, (\sigma, \sigma') \in \mathcal{R}^{\mathsf{O}}$, $M, \sigma' \models \varphi$
  - *(sf9)* $M, \sigma \models \mathsf{CAP}(\alpha)$ *iff* $\sigma \in c^A(\alpha)$
  - *(sf10)* $M, \sigma \models \mathsf{RES}(\alpha)$ *iff* $\sigma \in r^A(\alpha)$
- *satisfaction for path-formulas:*
  - *(pf1)* $M, \pi_{\theta_\sigma} \models \mathbf{X}\varphi$ *iff* $M, \pi_{\theta_\sigma}[1] \models \varphi$
  - *(p2f)* $M, \pi_{\theta_\sigma} \models \varphi_1 \mathbf{U} \varphi_2$ *iff* $\exists\, k \geq 0$ *such that* $M, \pi_{\theta_\sigma}[k] \models \varphi_2$ *and* $\forall j,\ 0 \leq j < k,\ M, \pi_{\theta_\sigma}[j] \models \varphi_1$

If, in all $\mathcal{EBDI}$-models M and situations $\sigma \in \Sigma$, $M, \sigma \models \varphi$, then φ is valid. If it is the case that $M, \sigma \models \varphi$ only for some M and σ, then φ is satisfiable in model M and situation σ.

Properties of time The temporal layer of $\mathcal{EBDI}$ corresponds to **CTL** logic [10]. Therefore, we have the path operators $\mathbf{A}\psi$ and $\mathbf{E}\psi$, which assert that ψ holds over all paths, and at least in one of them, respectively. For reasoning about the properties of a particular path, we have the modal operators $\varphi_1 \mathbf{U} \varphi_2$ and $\mathbf{X}\varphi$. These express the conditions that φ_1 holds until φ_2 holds, and φ holds at the next state of the path. When combined with the path quantifying operators, we have a lot of expressiveness to reason about the behaviour of an agent which, in each state of its computation, is faced with the task of deciding which path to follow.

As in **CTL**, the following axioms verify:

$(ctl1)$ $\mathbf{AG}(\varphi \rightarrow \psi) \rightarrow (\mathbf{EX}\varphi \rightarrow \mathbf{EX}\psi)$

$(ctl2)$ $\mathbf{EX}\top \wedge \mathbf{AX}\top$

$(ctl3)$ $\mathbf{E}(\varphi \mathbf{U} \psi) \leftrightarrow \psi \vee (\varphi \wedge \mathbf{EXE}(\varphi \mathbf{U} \psi))$

$(ctl4)$ $\mathbf{A}(\varphi \mathbf{U} \psi) \leftrightarrow \psi \vee (\varphi \wedge \mathbf{AXA}(\varphi \mathbf{U} \psi))$

$(ctl5)$ $\mathbf{AG}(\varphi \rightarrow (\neg\psi \rightarrow \mathbf{EX}\varphi)) \rightarrow (\varphi \rightarrow \neg\mathbf{A}(\varphi \mathbf{U} \psi))$

$(ctl6)$ $\mathbf{AG}(\varphi \rightarrow (\neg\psi \rightarrow \mathbf{EX}\varphi)) \rightarrow (\varphi \rightarrow \neg\mathbf{AF}\psi))$

$(ctl7)$ $\mathbf{AG}(\varphi \rightarrow (\neg\psi \rightarrow (\gamma \wedge \mathbf{AX}\varphi))) \rightarrow (\varphi \rightarrow \neg\mathbf{E}(\gamma \mathbf{U} \psi))$

$(ctl8)$ $\mathbf{AG}(\varphi \rightarrow (\neg\psi \rightarrow \mathbf{AX}\varphi)) \rightarrow (\varphi \rightarrow \neg\mathbf{EF}\psi))$

The set containing only the above axioms is denoted by CTL.

Properties of regular actions Regular actions provide high-level constructs which are suited to describe actions which an agent can execute upon its environment.

$\mathcal{EBDI}$ is based in PDL [12] and therefore verifies the following axioms

$(a1)\ \langle\alpha;\beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\langle\beta\rangle\varphi$
$(a2)\ \langle\alpha+\beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\varphi \vee \langle\beta\rangle\varphi$
$(a3)\ \langle\alpha^*\rangle\varphi \rightarrow \varphi \wedge \langle\alpha\rangle\varphi$
$(a4)\ \langle\alpha^*\rangle\varphi \rightarrow \langle\alpha\rangle\langle\alpha^*\rangle\varphi$
$(a5)\ \varphi \wedge \langle\alpha^*\rangle(\varphi \rightarrow \langle\alpha\rangle\varphi) \rightarrow \langle\alpha^*\rangle\varphi$

The set containing only the above axioms is denoted as PDL.

Relations between time and actions Time and action interact with each other in the following sense: if after executing successfully a particular action α the proposition φ holds, then it is also true that there exists in the future a state where the proposition φ also holds. However, the inverse case in not true, since φ may hold as the result of executing an action β, different from α. Formally, we have the following two axioms:

$(ta1)\ \langle a\rangle\varphi \rightarrow \mathbf{EX}\varphi$
$(ta2)\ \langle\alpha\rangle\varphi \rightarrow \mathbf{EF}\varphi$

As an example, consider the following scenarios:

- the agent, after driving a vehicle at high-speed, was not able to stop at time and crashed. Formally

  $\langle \mathsf{KeepHighSpeed}^*\rangle CrashedCar$

- the agent, after driving a vehicle for some time crashed it.
  $\mathbf{EF}(CrashedCar)$

It is perfectly acceptable that the crashed car after some high-speed driving imply that the car will be crashed in the future. However, the vehicle being crashed in the future does not necessarily imply that the cause was driving at high speed.

BDI layer For beliefs we use the KD-45 axiom system and the axiom system KD for both desires and intentions, as in [10]. Therefore, the set BEL_{KD45} for beliefs contains the following axioms:

$(belK)$ $\mathsf{BEL}(\varphi \rightarrow \psi) \rightarrow (\mathsf{BEL}(\varphi) \rightarrow \mathsf{BEL}(\psi))$
$(belD)$ $\mathsf{BEL}(\varphi) \rightarrow \neg\mathsf{BEL}(\neg\varphi)$
$(bel4)$ $\mathsf{BEL}(\varphi) \rightarrow \mathsf{BEL}(\mathsf{BEL}(\varphi))$
$(bel5)$ $\neg\mathsf{BEL}(\varphi) \rightarrow \mathsf{BEL}(\neg\mathsf{BEL}(\varphi))$

while DES_{KD} and INT_{KD} sets, for desires and intentions, contain respectively the first two and second two of the following axioms:

$(desK)$ $\mathsf{DES}(\varphi \rightarrow \psi) \rightarrow (\mathsf{DES}(\varphi) \rightarrow \mathsf{DES}(\psi))$
$(desD)$ $\mathsf{DES}(\varphi) \rightarrow \neg\mathsf{DES}(\neg\varphi)$

$(intK)$ $\mathsf{INT}(\varphi \rightarrow \psi) \rightarrow (\mathsf{INT}(\varphi) \rightarrow \mathsf{INT}(\psi))$
$(intD)$ $\mathsf{INT}(\varphi) \rightarrow \neg\mathsf{INT}(\neg\varphi)$

Capabilities, resources and actions Informally, we can see both the capabilities and resources as prerequisites for successful action-execution.

Resources and capabilities are defined in the Emotional-BDI model as follows:

Resources: these are physical/virtual means which the may drawn in order to make the agent capable of executing actions. If the resources for executing some action α do not exist, the action's success may be at stake.

Capabilities: these are abstract means which the agent has to change the environment in some way, thus resembling to abstract plans of action. In fact, we can consider the set of capabilities as a dynamic set of plans which the agent has available to decide what to do in each of its execution states.

In $\mathcal{EBDI}$, the axioms which characterise these concepts are

$(f1)$ $\mathsf{f}(\alpha;\beta) \rightarrow \mathsf{f}(\alpha) \wedge \langle\alpha\rangle\mathsf{f}(\beta)$
$(f2)$ $\mathsf{f}(\alpha+\beta) \rightarrow \mathsf{f}(\alpha) \vee \mathsf{f}(\beta)$
$(f3)$ $\mathsf{f}(\alpha^*) \rightarrow \mathsf{f}(\alpha) \wedge \langle\alpha\rangle\mathsf{f}(\alpha^*)$
$(f4)$ $\mathsf{f}(\alpha) \wedge \langle\alpha^*\rangle(\mathsf{f}(\alpha) \rightarrow \langle\alpha\rangle\mathsf{f}(\alpha)) \rightarrow \mathsf{f}(\alpha^*)$

with $\mathsf{f} \in \{\mathsf{CAP}, \mathsf{RES}\}$, and define the sets CAP and RES.

Since agents live in complex and highly dynamic environments, the information they capture may contain too much noise. However, it is in this noisy information the agent relies on, and which affects the information the agent has about its own means. This is what we call *effective capabilities* [4, 2], which are the (possibly wrong) beliefs about capabilities and resources. Formally it is expressed as $\mathsf{EffCap}(\alpha) \equiv \mathsf{BEL}(\mathsf{CAP}(\alpha)) \wedge \mathsf{BEL}(\mathsf{RES}(\alpha))$. This allows us to model acceptable facts such as $\mathsf{EffCap}(\alpha) \wedge \langle\alpha\rangle\bot$, which expresses the fact that, based on sufficiently wrong information about resources and capabilities, an agent may not succeed in performing an action, as expected.

On the other hand, if we know that an action was successfully executed, then it is true that the agent had effective capabilities which lead him to execute the action. Formally this is written as $\langle\alpha\rangle\top \rightarrow \mathsf{EffCap}(\alpha)$.

Fear Fear, in $\mathcal{EBDI}$, is explicitly referred by the modal operator FEAR. This operator should be read as *the agent fears that* φ *verifies.*

For fear we require only the Kripke-axiom

$$\mathsf{FEAR}(\varphi \to \psi) \to (\mathsf{FEAR}(\varphi) \to \mathsf{FEAR}(\psi))$$

to verify, and the set containing only this axiom is denoted by $FEAR_K$.

Fundamental Desires Fundamental desires are special desires which are vital desires of the agent, or desires which cannot be failed to achieve, in any condition, since may put in cause the agent's own existence. Fundamental desires must always be true and the agent must always do its best to maintain them valid.

The set of axioms which describe FDES are the following

$(fdesK)$ $\mathsf{FDES}(\varphi \to \psi) \to (\mathsf{FDES}(\varphi) \to \mathsf{FDES}(\psi))$
$(fdesT)$ $\mathsf{FDES}(\varphi) \to \varphi$
$(fdesD)$ $\mathsf{FDES}(\varphi) \to \neg\mathsf{FDES}(\neg\varphi)$

and we denote this set by $FDES_{KDT}$. This operator was introduced to facilitate the specification of triggering conditions for fear.

The basic Emotional-BDI system Now that all the modal operators were characterised, we are in conditions to define the simplest of Emotional-BDI agents. This is called an *basic* Emotional-BDI agent, and formally is defined by the union of all the sets describing each of the modal operators of $\mathcal{EBDI}$.

Definition 10. *The basic Emotional-BDI system is characterised by the union of the following sets of axioms:*

1. *the set of all propositional tautologies*
2. *the time axiom set* CTL
3. *the action axiom set* PDL
4. *the belief axiom set* BEL_{KD45}
5. *the desire axiom set* DES_{KD}
6. *the intention axiom set* INT_{KD}
7. *the capabilities axiom set* CAP
8. *the resources axiom set* RES
9. *the fear axiom set* $FEAR_K$
10. *the fundamental desire axiom set* $FDES_{KDT}$

Any other system to specify an agent in $\mathcal{EBDI}$ must extend this system. One such case is going to be presented in Section4.

4 Modelling Fear

Fear is one of the most studied emotions.The triggering of this emotion is usually associated with the fact that one or more fundamental desires are at stake.

However, this triggering does not always occur in a similar way in all agents. For instance, two different persons can fear the presence of a rattle-snake with completely distinct intensities. It is based on this differences that we can model different classes of Emotional-BDI agents. In this paper we introduce the *fearful agent*, which acts always in a very self-preservation way.

4.1 Threats and unpleasant facts

Negative emotions like fear occur when unwanted or uncontrollable facts or events are present on the environment. Here we consider only *threats* and *unpleasant facts*.

Threats represent facts or events occurring in the environment which directly affect one or more fundamental desires of the agent, putting at stake its self preservation. This threats may have different *weights*, which define how bad they may be for the agent. Here we consider the following:

Dangerous: a threat is dangerous when the agent believes that some condition ψ leads inevitably to the falsity of a fundamental desire φ, and also believes that ψ will also be inevitably true in the future.

$$\mathsf{DangerousThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{AF}\psi)$$

Serious: a threat is serious if the same conditions of a dangerous one hold, except that the agent believes that ψ may eventually be true in the future, and not always.

$$\mathsf{SeriousThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{EF}\psi)$$

Possible: a threat is possible if the fundamental desire continues at stake, but the agent believes that it may hold only in the future.

$$\mathsf{PossibleSeriousThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{EF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{EF}\psi)$$

Unpleasant facts represent facts or events which put one or more desires in risk of non-achievement. The agent may exhibit distinct behaviour towards such unpleasant fact, for protecting its desires. Here we consider the following:

Highly Unpleasant: something becomes highly unpleasant if the agent believes that the source of the unpleasantness will always occur in the future and will always in the future put in cause some desire.

$$\mathsf{HighlyUnpleasant}(\psi, \varphi) \equiv \mathsf{DES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{AF}\psi)$$

Strongly Unpleasant: something becomes strongly unpleasant if the agent believes that the source of the unpleasantness will eventually occur in the future and will always in the future put in cause some desire.

$$\mathsf{StronglyUnpleasant}(\psi, \varphi) \equiv \mathsf{DES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{EF}\psi)$$

Possibly Unpleasant: something becomes possibly unpleasant if the agent believes that the source of the unpleasantness will eventually occur in the future and, in the case of occurring, maybe it will put in cause some desire.

$$\mathsf{PossiblyUnpleasant}(\psi, \varphi) \equiv \mathsf{DES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{EF}(\neg\varphi)) \wedge \mathsf{BEL}(\mathbf{EF}\psi)$$

These concepts will be used in what follows to model the triggering of fear and to show that special processing strategies may be applied when facing certain conditions.

4.2 Special atomic actions

Based on the literature [4, 13], we define the following special purpose actions, which represent specific behaviour exhibited by the agent under certain conditions.

Self-preservation: the self-preservation behaviour is activated when the agent is fearing the failure of some of its fundamental desires. We can see this as atomic action which mainly reacts to threats in a self-protective way. In $\mathcal{EBDI}$, this special action is represented by $\mathsf{spreserv}$.

Motivated Processing Strategy: this processing strategy is employed by the agent when some desire which directs its behaviour must be maintained but may be at risk. This strategy is computationally intensive, as its should produce complex data-structures for preserving desires. In $\mathcal{EBDI}$, this kind of processing is abstracted into the specialized atomic action mpsdel.

Direct Access Strategy: this processing strategy relies on the use of fixed pre-existing structures/knowledge. It is the simplest strategy and corresponds to a minimization of the computational effort and to fast solutions. In $\mathcal{EBDI}$, this kind of processing is abstracted into the specialized atomic action dasdel.

Considering the above actions as being atomic actions is of course a big abstraction to the complexity of Emotional-BDI agents. These actions are usually complex planning and revision strategies.

4.3 Specifying a fearful agent

We will now introduce the *fearful* Emotional-BDI system, but first we are going to present the intuition behind it.

A fearful agent is an agent which exhibits a very careful behaviour, considering any threat as a fear factor, and considering all uncomfortable events at the same level. Being careful, the agent also acts towards solving threats and unpleasant facts with the best of its means, and even if the means for the best action are not available, the agent always tries to put itself on a safe, or self-preservation condition.

Let the following equivalences be defined

$$\begin{aligned}\mathsf{AnyThreat}(\psi,\varphi) &\equiv \mathsf{DangerousThreat}(\psi,\varphi) \vee \mathsf{SeriousThreat}(\psi,\varphi)\\ &\quad \vee \mathsf{PossibleSeriousThreat}(\psi,\varphi)\\ \mathsf{AnyUnpleasant}(\psi,\varphi) &\equiv \mathsf{HighlyUnpleasant}(\psi,\varphi) \vee \mathsf{StronglyUnpleasant}(\psi,\varphi)\\ &\quad \vee \mathsf{PossiblyUnpleasant}(\psi,\varphi)\end{aligned}$$

We define a fearful Emotional-BDI system as follows.

Definition 11. *Let φ, ψ be well-formed formulae and $\alpha \in A_{\mathsf{Ra}}$ be a regular action. The specification of fearful Emotional-BDI agent is given by all the axioms of an Emotional-BDI agent, plus the following new axioms:*

- *any threat to a fundamental desire φ makes the agent fear $\neg\varphi$*

  $$\mathsf{AnyThreat}(\psi,\varphi) \rightarrow \mathsf{FEAR}(\neg\varphi)$$

- *any threat ψ to a fundamental desire φ is also a fear of the agent*

  $$\mathsf{AnyThreat}(\psi,\varphi) \rightarrow \mathsf{FEAR}(\psi)$$

- *if the agent after deliberating, using the motivate processing strategy, believes that either ψ will not verify or that φ and ψ are compatible, then he considers ψ as an unpleasant fact for the achievement of φ*

  $$\langle \mathsf{mpsdel} \rangle ((\mathsf{BEL}(\neg\psi) \vee \mathsf{BEL}(\psi \wedge \varphi)) \wedge \mathsf{DES}(\varphi)) \rightarrow \mathsf{AnyUnpleasant}(\psi,\varphi)$$

- *if the agent is threatened and has no effective resources for executing a deliberation based on direct access strategies (which could bring good solutions for avoiding the threat), the agent can execute a self preservation action*

  $$\mathsf{AnyThreat}(\psi,\varphi) \wedge \neg\mathsf{EffCap}(\mathsf{dasdel}) \rightarrow \langle \mathsf{spreserv} \rangle \top$$

- *if the agent has effective capabiliities, it executes instead the direct access strategy based deliberation*

  $$\mathsf{AnyThreat}(\psi,\varphi) \wedge \mathsf{EffCap}(\mathsf{dasdel}) \rightarrow \langle \mathsf{dasdel} \rangle \top$$

It is important to stress that the presented fearful Emotional-BDI agent is not the unique extension to the basic Emotional-BDI agent. This system only characterises agents whose fear is triggered by any threat, and that reconsider its desires at the first unpleasant fact that interferes with them. If, for instance, we substituted $\mathsf{AnyThreat}(\psi,\varphi) \rightarrow \mathsf{FEAR}(\neg\varphi)$ by $\mathsf{DangerousThreat}(\psi,\varphi) \rightarrow \mathsf{FEAR}(\neg\varphi)$, we would be specifying Emotional-BDI agent which only fear dangerous threats.

5 Related work

The subject of formally modelling emotional agents was already addressed by J.J. Meyer in [14]. In his work, Meyer uses the **KARO** framework and imposes conditions on the structure where **KARO** is interpreted, so that the triggering of emotions (happiness, sadness, anger and fear) and their effects on the behaviour of the agent are conveniently defined.

Work was also done in introducing the notion of capability in Rao & Georgeff's $\mathsf{BDI}_{\mathbf{CTL}}$ logic. This work was presented in [11] but do not explicitly refer actions. It is only considered as the ability to rationally act towards the achievement of desires.

6 Conclusions and future work

In this paper we presented $\mathcal{EBDI}$ logic, a logic developed for modelling Emotional-BDI agents. We presented its syntax, semantics and showed its expressiveness by introducing the notions of threat, unpleasant fact and use them to model a class of Emotional-BDI agents which we call fearful agents.

Our approach was based in $\mathsf{BDI}_{\mathbf{CTL}}$, whose ideas will continue to be followed by us in our present and future work, which in particular concerns the development of complete deductive systems for $\mathcal{EBDI}$. Therefore, we want to obtain decision procedures for testing the satisfiability and validity of $\mathcal{EBDI}$-formulae. We are also interested in providing different Emotional-BDI systems reflecting other behaviour which Emotional-BDI agent can exhibit.

References

1. Bratman, M.E., Israel, D., Pollack, M.E.: Plans and resource-bounded practical reasoning. Computational Intelligence **4** (1988) 349–355
2. David Pereira, Eugénio Oliveira, N.M., Sarmento, L.: Towards an architecture for emotional bdi agents. In Carlos Bento, A.C., Dias, G., eds.: EPIA05 – 12^{th} Portuguese Conference on Artificial Intelligence, Universidade da Beira Interior, IEEE (2005) 40–46 ISBN 0-7803-9365-1.
3. Damasio, A.R.: Descartes' Error: Emotion, Reason and the Human Brain. Grosset/Putnam (1994)
4. Oliveira, E., Sarmento, L.: Emotional advantage for adaptability and autonomy. In: AAMAS. (2003) 305–312
5. Georgeff, M.P., Pell, B., Pollack, M.E., Tambe, M., Wooldridge, M.: The belief-desire-intention model of agency. In Müller, J.P., Singh, M.P., Rao, A.S., eds.: ATAL. Volume 1555 of Lecture Notes in Computer Science., Springer (1998) 1–10
6. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a BDI-architecture. In Allen, J., Fikes, R., Sandewall, E., eds.: Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91), Morgan Kaufmann publishers Inc.: San Mateo, CA, USA (1991) 473–484
7. van der Hoek, W., van Linder, B., Meyer, J.J.C.: A logic of capabilities. In Nerode, A., Matiyasevich, Y., eds.: LFCS. Volume 813 of Lecture Notes in Computer Science., Springer (1994) 366–378

8. Schmidt, R.A., Tishkovsky, D., Hustadt, U.: Interactions between knowledge, action and commitment within agent dynamic logic. Studia Logica **78**(3) (2004) 381–415
9. Schild, K.: On the relationship between bdi logics and standard logics of concurrency. Autonomous Agents and Multi-Agent Systems **3**(3) (2000) 259–283
10. Rao, A.S., Georgeff, M.P.: Decision procedures for bdi logics. J. Log. Comput. **8**(3) (1998) 293–342
11. Padgham, L., Lambrix, P.: Formalisations of capabilities for bdi-agents. Autonomous Agents and Multi-Agent Systems (10) (2005) 249–271
12. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. MIT Press (2000) HAR d 00:1 1.Ex.
13. Sarmento, L.: An emotion-based agent architecture. Master's thesis, Faculdade de Ciências da Universidade do Porto (2004)
14. Meyer, J.J.C.: Reasoning about emotional agents. In de Mántaras, R.L., Saitta, L., eds.: ECAI, IOS Press (2004) 129–133