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 $\mathcal{E}_{\mathsf{BDI}}$ 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 $\mathcal{E}_{\mathsf{BDI}}$ 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 an additional emotional component [2]. This component produces data which will bound the BDI interaction by imposing some of the 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 recent works of Oliveira & Sarmento's about an emotional agent architecture [4], although adapted to fit in the original BDI architecture [1, 5, 6].

In this paper we introduce $\mathcal{E}_{\mathsf{BDI}}$, a multi-modal logic for specifying Emotional-BDI agents. We define the various axioms which properly characterise each of the modal operators of $\mathcal{E}_{\mathsf{BDI}}$ 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 current work; in Section 3 we introduce the logic $\mathcal{E}_{\mathsf{BDI}}$ and define its syntax and semantics, together with the 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 related work 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 describes 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 $\mathcal{E}_{\mathsf{BDI}}$'s syntax.

3 The Logic $\mathcal{E}_{\mathsf{BDI}}$

We will now introduce the logic $\mathcal{E}_{\mathsf{BDI}}$. 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 $\mathcal{E}_{\mathsf{BDI}}$ is a two dimensional structure introduced by Schild [9], which is a simplified approach to Rao & Goergeff's $\mathsf{BDI}_{\mathbf{CTL}}$ [10] semantics. One dimension is a set of possible worlds that corresponds to the different prescpectives of the agent, such as its beliefs, desires, etc. The other is a set of temporal states which describe the temporal evolution of the agent. We call a pair $\langle world, temporal\_state \rangle$ a *situation*.

In $\mathcal{E}_{\mathsf{BDI}}$, as in the **KARO** framework, we consider explicit complex actions. 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 a set of regular rules. Actions are a labelling of the temporal structure underlying $\mathcal{E}_{\mathsf{BDI}}$.

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 $\mathcal{E}_{\mathsf{BDI}}$. 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 $\mathcal{E}_{\mathsf{BDI}}$ 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 temporal 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 $\mathcal{E}_{\mathsf{BDI}}$* well-formed formulas *is 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 $\varphi \to \psi$ 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 $\neg(\varphi \wedge \neg\psi)$, 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 α^n, with $n \geq 0$, 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 $\mathcal{E}_{\mathsf{BDI}}$. We start by defining the notion of situation.

Definition 2. *Given a 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 $\sigma = \langle w_i, t_j\rangle$, with $i \geq 0$ and $j \geq 0$. The set of situations is denoted by 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 in the situation $\langle belief, t\rangle$, 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 $\mathcal{R}_T$ and the other is a *action execution* relation that associates to each element of $\mathcal{R}_T$ an atomic action.

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* $w_i = w_k$.

Only imposing that $\mathcal{R}_T$ *is only serial, and not a total (linear) order, leads to a branching-time structure.*

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 an action execution relation* $\mathcal{R}_{a_i}$*, such that:*

1. $\mathcal{R}_{a_i} \in \mathcal{R}_T$*;*
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:*

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

The main interest behind using both approaches 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 a situation 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 $\mathcal{E}_{\mathsf{BDI}}$, as in $\mathsf{BDI}_{\mathbf{CTL}}$, the former are analysed along a path (a time branch) and the second in a particular situation. In $\mathcal{E}_{\mathsf{BDI}}$, 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* $\forall i \geq 0$, $(\sigma_i, \sigma_{i+1}) \in \mathcal{R}_T$*. The* k^{th} *element of a path* $\pi\sigma$ *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 world component of any situation $\langle \mathit{world}, \mathit{temporal\_state} \rangle$. The accessibility relations which establish this relationship are the ones which are going to be used for modelling the mental states of the agent. These relations are denoted by $\mathcal{R}^{\mathsf{O}}$, with O belonging a set of modal operators and that must respect the following condition: if $(\langle w_i, t_j\rangle, \langle w_k, t_l\rangle) \in \mathcal{R}^{\mathsf{O}}$ then $t_j = t_l$.

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, i.e., 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 $\mathrm{v}_f(a_i)$ which establishes a subset of Σ where the agent has capabilities/resources to execute atomic actions $a_i \in A_{\mathsf{At}}$, resources and capabilities are interpreted by similar functions. Therefore, we inductively define them in a function f, with $f \in \{c, r\}$, such that:*

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

The interpretation of $\mathcal{E}_{\mathsf{BDI}}$-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 $\mathcal{E}_{\mathsf{BDI}}$-*model *as a tuple*

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

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;*
- *$\mathrm{v}_p, \mathrm{v}_c$ and 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 $\mathcal{E}_{\mathsf{BDI}}$ is given by the following definition.

Definition 9. *Let M be an $\mathcal{E}_{\mathsf{BDI}}$-model. The* satisfiability *of a $\mathcal{E}_{\mathsf{BDI}}$-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* $p \in \mathrm{v}_p(\sigma)$
  - *(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* $\exists \pi\sigma$ *such that* $M, \pi\sigma \models \psi$
  - *(sf5)* $M, \sigma \models \mathbf{A}\psi$ *iff* $\forall \pi\sigma,\ M, \pi\sigma \models \psi$
  - *(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)* $M, \pi\sigma \models \mathbf{X}\varphi$ *iff* $M, \pi\sigma[1] \models \varphi$
  - *(p2f)* $M, \pi\sigma \models \varphi_1 \mathbf{U} \varphi_2$ *iff* $\exists\ k \geq 0$ *such that* $M, \pi\sigma[k] \models \varphi_2$ *and* $\forall j,\ 0 \leq j < k,\ M, \pi\sigma[j] \models \varphi_1$

If $M, \sigma \models \varphi$ in all $\mathcal{E}_{\mathsf{BDI}}$-models M and situations $\sigma \in \Sigma$, 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 M and situation σ.

Properties of time The temporal layer of $\mathcal{E}_{\mathsf{BDI}}$ corresponds to **CTL** logic [10]. Therefore, we have the formulas $\mathbf{A}\psi$ 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 formulas $\varphi_1 \mathbf{U} \varphi_2$ 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. As in **CTL**, the following axioms verify:

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

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

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

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

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

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

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

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

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.

$\mathcal{E}_{\mathsf{BDI}}$ is based in PDL [12] and therefore the following axioms verify

$(a1)$ $\langle \alpha;\beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\langle\beta\rangle\varphi$

$(a2)$ $\langle \alpha+\beta\rangle\varphi \leftrightarrow \langle\alpha\rangle\varphi \vee \langle\beta\rangle\varphi$

$(a3)$ $\langle \alpha^*\rangle\varphi \rightarrow \varphi \vee \langle\alpha\rangle\langle\alpha^*\rangle\varphi$

$(a4)$ $\varphi \wedge \langle\alpha^*\rangle(\varphi \rightarrow \langle\alpha\rangle\varphi) \rightarrow \langle\alpha^*\rangle\varphi$

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

Lets now define some properties relating regular actions to temporal formulae.

Lemma 1. *Let M be a $\mathcal{E}_{\mathsf{BDI}}$-model and σ a situation. If $M,\sigma \models \langle\alpha^*\rangle\varphi$ then $M,\sigma \models \varphi \vee \langle\alpha^n\rangle\varphi$, for $n \in \mathbb{N}, n \geq 1$.*

Lemma 2. *Let M be a $\mathcal{E}_{\mathsf{BDI}}$-model and σ a situation. If $M,\sigma \models \langle\alpha^n\rangle\varphi$, for $n \geq 1$, then $M,\sigma \models \langle\alpha\rangle\mathbf{E}(\langle\alpha\rangle\top\mathbf{U}\varphi)$.*

Theorem 1. *Let M be a $\mathcal{E}_{\mathsf{BDI}}$-model and σ a situation. If $M,\sigma \models \langle\alpha^*\rangle\varphi$ then $M,\sigma \models \varphi \vee \langle\alpha\rangle\mathbf{E}(\langle\alpha\rangle\top\mathbf{U}\varphi)$.*

Relations between time and actions Time and action interact with each other in the following sense: if after successfully executing 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 is not true, since φ may hold as the result of executing an action β, different from α. Formally, we have the following two axioms:

Theorem 2. *Let M be an $\mathcal{E}_{\mathsf{BDI}}$-model, and $\sigma \in \Sigma_M$. Then the following formulae are theorems of $\mathcal{E}_{\mathsf{BDI}}$:*

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

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

As an example, consider the following scenarios:

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

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

- 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:

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

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

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

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

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 may be 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 $\mathcal{E}_{\mathsf{BDI}}$, the axioms which characterise these concepts are

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

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

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 $\langle\alpha\rangle\top \rightarrow \mathsf{EffCap}(\alpha)$.

Theorem 3. *Let M be a $\mathcal{E}_{\mathsf{BDI}}$-model, and σ a situation. Then, if $M, \sigma \models \mathsf{CAP}(\alpha^*)$ then $M, \sigma \models \mathbf{E}((\mathsf{CAP}(\alpha) \wedge \langle\alpha\rangle\mathsf{CAP}(\alpha))\mathbf{U}\top)$.*

Theorem 4. *Let M be a $\mathcal{E}_{\mathsf{BDI}}$-model, and σ a situation. Then, if $M, \sigma \models \mathsf{RES}(\alpha^*)$ then $M, \sigma \models \mathbf{E}((\mathsf{RES}(\alpha) \wedge \langle\alpha\rangle\mathsf{RES}(\alpha))\mathbf{U}\top)$.*

Fear Fear, in $\mathcal{E}_{\mathsf{BDI}}$, 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

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

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 danger the agent's own existence. Fundamental desires should 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

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

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 Emotional-BDI agents. This is called the *basic* Emotional-BDI agent.

Definition 10. *A basic Emotional-BDI system is a set of formulae which is contain 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 $FDES_{KD}$

and that are closed under the inference rules of modus ponens $\varphi, \varphi \rightarrow \psi \Rightarrow \psi$ *and the necessitation rule* $\vdash \varphi \ \Rightarrow \vdash \Box\varphi$*, where* $\Box \in \{\mathsf{BEL}, \mathsf{DES}, \mathsf{INT}, \mathsf{FDES}, \mathsf{FEAR}, \mathbf{AG}, [\alpha]\}$*, with* α *being a regular action.*

Any other system to specify an agent in $\mathcal{E}_{\mathsf{BDI}}$ must extend this system. One such case is going to be presented in Section 4.

4 Modelling Fear

Agents are affected by fear in different ways, depending on how their internal representations differentiate between what are dangerous situations or non-dangerous situations. These differences of fear reactions have a direct impact on how agents may react in distinct ways with respect to some situation. For instance, a civilian may elicit fear about get shot just by earing some fire shots, while a policeman or a soldier element may get only alert, due to its everyday contact with highly dangerous situations.

4.1 Threats

Negative emotions like fear are generally elicited when some possibly dangerous conditions of the environment (or generated by the agent) put at stake one of the agent's fundamental goals. This may also put in cause the agent's own self-preservation. Here, these conditions are called *threats*.

Threats can be scaled in terms of their dangerousness and time occurrence. By this we mean that there are threats which are more dangerous than others, and threats which already are present on the environment and others which most likely will end up by occuring in the environment.

Current threats: the source of the threat is occurring now, and the agent has information about the fact that the existance of such source may put at stake its fundamental goals.

- $\mathsf{VeryDangerousCThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \neg\varphi) \wedge \psi$
- $\mathsf{DangerousCThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \psi$
- $\mathsf{CThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{EF}(\neg\varphi)) \wedge \psi$

Future threats: the source of the threat will eventually occur in the future.

- $\mathsf{VeryDangerousPThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \neg\varphi) \wedge \mathbf{AF}\psi$
- $\mathsf{DangerousPThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{AF}(\neg\varphi)) \wedge \mathbf{AF}\psi$
- $\mathsf{PThreat}(\psi, \varphi) \equiv \mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\psi \rightarrow \mathbf{EF}(\neg\varphi)) \wedge \mathbf{AF}\psi$

In this paper, we formally model these classes of agents in order to show that our logic is expressive enough to model different kinds of agents, which generally react differently to distinct types of threats.

Now, a general threat – being it current or possible in the future – is any threat, with any amount of associated danger. Formally,

$$\begin{aligned}\mathsf{AnyCThreat}(\psi,\varphi) \equiv\ & \mathsf{VeryDangerousCThreat}(\psi,\varphi) \vee \mathsf{DangerousCThreat}(\psi,\varphi) \vee \\ & \mathsf{CThreat}(\psi,\varphi) \\ \mathsf{AnyPThreat}(\psi,\varphi) \equiv\ & \mathsf{VeryDangerousPThreat}(\psi,\varphi) \vee \mathsf{DangerousPThreat}(\psi,\varphi) \vee \\ & \mathsf{PThreat}(\psi,\varphi)\end{aligned}$$

4.2 Special atomic actions

Based on the literature [4, 13], we will introduce a set of special purpose actions, which represent specific behaviour exhibited by the agent under certain emotional conditions. These actions are information processing strategies identified in humans [14] and which are applied by them for obtaining solutions under specific emotional states. Here we will only present the strategies which had been identified as being activated under fear conditions.

Besides these strategies, we also introduce an abstract *self-preservation* action, whose meaning is the reactive character of an agent when the urgency for avoiding a dangerous situations is so great that none of the other processing strategies will provide good solutions in an acceptable time.

The set of special actions we define is:

1. *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 $\mathcal{E}_{\mathsf{BDI}}$, this special action is represented by $\mathsf{selfpreservation}$.
2. *Direct Access:* this processing strategy relies on the use of fixed pre-existing structures/knowledge. It is the simplest strategy and corresponds to a minimisation of the computational effort and to fast solutions. In $\mathcal{E}_{\mathsf{BDI}}$, this kind of processing is abstracted into the specialised atomic action das.
3. *Motivated Processing:* 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 $\mathcal{E}_{\mathsf{BDI}}$, this kind of processing is abstracted into the specialised atomic action mps.
4. *Substantive Processing:* this is considered the most complex information processing strategy and is usually applied to obtain possible solutions for situations which require large amount of computational effort for obtaining complex plans. It is applied when there are enough resources and capabilities and not too much urgency on find a solution. In $\mathcal{E}_{\mathsf{BDI}}$, this kind of processing is denoted by the atomic action sps.

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 present a formal specification of a what we consider a *fearful Emotional-BDI agent*. Informally, a fearful Emotional-BDI agent describes a class of software agents which elicit fear in all the situations where threats (or possible threats) are detected, not distinguishing between really dangerous threats or only light or possible threats. However, the temporal characteristics of the threats are taken in account by the agent, which fears their proximity. Based on what are the fears of the agent, it will employ distinct deliberation strategies studied in the literature [14], which require distinct levels of resources and capabilities, depending on what kind of urgent situations they are to be applied to.

The formal specification of fearful Emotional-BDI agents will be done in two parts:

- *eliciting conditions*, which are $\mathcal{E}_{\mathsf{BDI}}$ formulae which explicitly define in which situations the agent elicits fear about propositions;
- *behaviour effect*, which are $\mathcal{E}_{\mathsf{BDI}}$ formulae which state what kind of behaviour is exhibited by the agent in order to avoid the fears it has elicited and are still present in the agent's internal state.

The elicitation of fear The agent elicits fear about some proposition if that proposition describes a threatening situation to one of its fundamental desires.

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

If the threat is still to occur, the agent will fear not the threat itself, but its future occurrence.

$$\mathsf{AnyPThreat}(\psi, \varphi) \rightarrow \mathsf{FEAR}(\mathbf{AF}\psi)$$

Now, if the agent already has beliefs about how to achieve a certain fundamental desire (or on how to maintain it), the will fear situations where unexpected interruptions on the execution of the actions to achieve that occur. In a first case, if the agent detects that it doesn't have effective capabilities to successfully accomplish the action, it will fear for that lack of effective capabilities.

$$\mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\langle \alpha; \beta \rangle \varphi) \rightarrow [\alpha](\neg\mathsf{EffCap}(\beta) \rightarrow \mathsf{FEAR}(\neg\mathsf{EffCap}(\beta)))$$

But the agent may only detect the fact that, even though it has effective resources to execute the rest of the action, the successful execution of that action will possibly lead to a non wanted falsity of the fundamental desire. In this case, the agent will fear for a successfully execution of the action

$$\mathsf{FDES}(\varphi) \wedge \mathsf{BEL}(\langle \alpha; \beta \rangle \varphi) \rightarrow [\alpha](\mathsf{BEL}([\beta]\neg\varphi) \rightarrow \mathsf{FEAR}(\langle \beta \rangle \top))$$

The effects of fear in behavior If the agent is present before a current threat and it does not believe that it will obtain a good solution using even the quickest and less computational requiring processing strategy before the threatened fundamental desire becomes false, it will execute the self-preservation action in order to, at least, guarantee its most basic safety condition

$$\mathsf{FEAR}(\psi) \wedge \mathsf{AnyCThreat}(\psi, \varphi) \wedge \mathsf{BEL}(\mathbf{A}(\neg\mathsf{EffCap}(\mathsf{das})\mathbf{U}\neg\varphi)) \rightarrow \langle\mathsf{selfpreservation}\rangle\top$$

However, if the agent believes it has effective capabilities to execute a direct processing strategy, and therefore obtain better solutions to avoid the threat, it will execute the direct processing instead of just safe-guarding itself

$$\mathsf{FEAR}(\psi) \wedge \mathsf{AnyCThreat}(\psi, \varphi) \wedge \mathsf{BEL}(\mathsf{EffCap}(\mathsf{das})) \rightarrow \langle\mathsf{das}\rangle\top$$

In the case of the threat is still to occur, the agent will employ either the motivated processing or substantive processing strategies, since they have still some time until the threat occurs and during this time they main obtain plans detailed enough to have a better guaranteed of avoiding the threat

$$\mathsf{FEAR}(\mathbf{AF}\psi) \wedge \mathsf{AnyPThreat}(\psi, \varphi) \wedge \mathsf{BEL}(\mathsf{EffCap}(\mathsf{mps} + \mathsf{sps})) \rightarrow \langle\mathsf{mps} + \mathsf{sps}\rangle\top$$

If the fear of the agent is elicited during the execution of one action supposed to achieve or maintain a fundamental desire, the agent must exhibit a behaviour which allow it to obtain an alternative action to fullfil the first action's goal

$$\mathsf{BEL}(\langle\alpha;\beta\rangle\varphi) \wedge \mathsf{FDES}(\varphi) \wedge [\alpha]((\mathsf{FEAR}(\neg\mathsf{EffCap}(\beta)) \vee \mathsf{FEAR}(\langle\beta\rangle\top)) \wedge \mathsf{EffCap}(\mathsf{das}))$$

$$\rightarrow [\alpha]\langle\mathsf{das}\rangle\top$$

If it does not has the effective capabilities to do it, the agent must self preserve itself before doing something else

$$\mathsf{BEL}(\langle\alpha;\beta\rangle\varphi) \wedge \mathsf{FDES}(\varphi) \wedge [\alpha]((\mathsf{FEAR}(\neg\mathsf{EffCap}(\beta)) \vee \mathsf{FEAR}(\langle\beta\rangle\top)) \wedge \neg\mathsf{EffCap}(\mathsf{das}))$$

$$\rightarrow [\alpha]\langle\mathsf{selfpreservation}\rangle\top$$

5 Related work

The subject of formally modelling emotional agents was already addressed by J.J. Meyer in [15]. 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 the syntax and semantics of $\mathcal{E}_{\mathsf{BDI}}$ logic, a logic developed for modelling Emotional-BDI agents. By introducing the notions of threat and unpleasant fact we have showed its expressiveness to model a class of Emotional-BDI agents which we called fearful agents.

Our approach was based in $\mathsf{BDI}_{\mathbf{CTL}}$ extended with explicit reference to actions, resources and capabilities. However, for satisfiability purposes, we can transform any $\mathcal{E}_{\mathsf{BDI}}$ formula into an $\mathsf{BDI}_{\mathbf{CTL}}$ formula. In this way, we can easily extend the decision procedures given for $\mathsf{BDI}_{\mathbf{CTL}}$ [10] to $\mathcal{E}_{\mathsf{BDI}}$.In particular, we can obtain the decidibility of the satisfiability problem of $\mathcal{E}_{\mathsf{BDI}}$ formulae, as well as the soundness and completness of the basic $\mathcal{E}_{\mathsf{BDI}}$ system, with respect to a class of models. This is part of our ongoing work. 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$^{\text{th}}$ 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. Gordan H. Bower, J.P.F.: Affect, memory and social cognition. In Marschark, M., ed.: Cognition and Emotion. Counterpoints: Cognition, Memory & Language. Oxford University Press (2000) 87–168
15. Meyer, J.J.C.: Reasoning about emotional agents. In de Mántaras, R.L., Saitta, L., eds.: ECAI, IOS Press (2004) 129–133