OCC's emotions: a formalization in a BDI logic

Carole Adam, Benoit Gaudou, Andreas Herzig, Dominique Longin

OCC's emotions: a formalization in a BDI logic*

Carole Adam, Benoit Gaudou, Andreas Herzig, and Dominique Longin

Université Paul Sabatier – IRIT/LILaC
118 route de Narbonne, F-31062 Toulouse CEDEX 9 (France)
{adam,gaudou,herzig,longin}@irit.fr

Abstract. Nowadays, more and more artificial agents integrate emotional abilities, for different purposes: expressivity, adaptability, believability... Designers mainly use Ortony et al.'s typology of emotions, that provides a formalization of twenty-two emotions based on psychological theories. But most of them restrain their agents to a few emotions among these twenty-two ones, and are more or less faithful to their definition. In this paper we propose to extend standard BDI (belief, desire, intention) logics to account for more emotions while trying to respect their definitions as exactly as possible.

1 Introduction

Recently, the agent community gets very interested in emotional artificial agents with enhanced abilities including expressivity, adaptability and believability. To cope with the increasing complexity of such agents, designers need rigorous formal models offering properties like genericity and verifiability.

Current computer models of emotions are mainly semi-formal or only manage a limited number of emotions, and are thus often specific to the context of use. However, Meyer [1] proposes a very formal model of four emotions, but as he states himself [1, p.11], his goal was not to "capture the informal psychological descriptions exactly (or as exact as possible)" but rather to describe what "makes sense for artificial agents". In this paper we provide a logical formalization of twenty emotions while staying as close as possible to one of the most cited psychological approaches, *viz.* that of Ortony, Clore, and Collins (OCC) [2]. Compared to the previous version of this work, we manage twelve more emotions, and we have modified the existing ones to be more faithful to OCC. These emotions are formalized inside a BDI modal logic (Belief, Desire, Intention), that has numerous advantages: widespread thus broadly studied, established results of verifiability and genericity, strong explicative power of the agent's behavior... Our architecture grounds on previous work [3]. We here omit the notions of choice and intention (that turned out to be avoidable), and add a

* A preliminary version of this work has been published in the ECAI worshop AITaMI'06. The authors would like to thank the AIMSA reviewers for their very useful comments.

probabilistic belief operator, as well as a refined notion of desire with its symmetric notion of "undesire", and deontic operators.

There exist several kinds of models of emotions. Discrete models (*e.g.* [5, 6]) are mainly descriptive. Dimensional models (*e.g.* [7]) are practical when aiming at describing the dynamics and expression of emotions (*e.g.* [8]) but not explicative of the triggering of emotions. Finally, cognitive models (*e.g.* [2, 9, 10]) assume that emotions are triggered by the cognitive evaluation of stimuli following some judgement criteria or *appraisal variables*. They are more normative than other models, and thus we can find them as a basis in many intelligent agent architectures. Some rare researchers prefer the complex theories from Lazarus (*e.g.* [11]) or Frijda (*e.g.* [12]), but most of them (*e.g.* [13]), including this paper, ground on the model of Ortony, Clore, and Collins (OCC).

The OCC typology has three branches, each of which corresponds to the appraisal of a different type of stimulus with respect to a particular appraisal variable, and related to particular mental attitudes. For example, the stimulus event "it is raining" is appraised as being undesirable w.r.t. the agent's goal of taking coffee on a terrace. These branches are then differentiated into several groups of emotion types with similar eliciting conditions.

Section 2 introduces our logical framework allowing to express the necessary intensity variables. Sections 3 and 4 detail the event-based and agent-based branches of the typology.

2 Logical framework

Our formal framework is based on the modal logic of belief, choice, time, and action of Herzig and Longin [3] which is a refinement of Cohen and Levesque's works [14]. We need neither choice nor intention (build from belief and choice), thus we do not use them. We extend this logic with modal probability operators defined by Herzig [4], as well as obligation and desirability operators.

Semantics. Let AGT be the set of agents, ACT the set of actions, $ATM = \{p, q...\}$ the set of atomic formulas. The set of complex formulas will be noted $FORM = \{\varphi, \psi...\}$. A possible-worlds semantics is used, and a model $\mathcal{M}$ is a triple $\langle W, V, \mathcal{R}\rangle$ where W is a set of possible worlds, V is a truth assignment which associates each world w with the set V_w of atomic propositions true in w, and $\mathcal{R}$ is a tuple of structures made up of:

– $\mathcal{A} : ACT \rightarrow (W \rightarrow 2^W)$ which associates each action $\alpha \in ACT$ and possible world $w \in W$ with the set $\mathcal{A}_\alpha(w)$ of possible worlds resulting from the execution of action α in w;

– $\mathcal{B} : AGT \rightarrow (W \rightarrow 2^W)$ which associates each agent $i \in AGT$ and possible world $w \in W$ with the set $\mathcal{B}_i(w)$ of possible worlds compatible with the beliefs of agent i in w. All these accessibility relations are serial;

– $\mathcal{P} : AGT \rightarrow (W \rightarrow 2^{2^W})$ which associates each agent $i \in AGT$ and possible world $w \in W$ with a set of sets of possible worlds $\mathcal{P}_i(w)$. Intuitively for $U \in \mathcal{P}_i(w)$, U contains more elements than its complement $\mathcal{B}_i \setminus U$;

– $\mathcal{G} : W \rightarrow 2^W$ which associates each possible world $w \in W$ with the set $\mathcal{G}(w)$ of possible worlds in the future of w. This relation is a linear order (reflexive, transitive and antisymmetric). $\mathcal{G} \supseteq \mathcal{A}_\alpha$ for every α;

– $\mathcal{L} : AGT \rightarrow (W \rightarrow 2^W)$ (resp. $\mathcal{L} : AGT \rightarrow (W \rightarrow 2^W)$) which associates each agent $i \in AGT$ and possible world $w \in W$ with the set $\mathcal{L}_i(w)$ (resp. $\mathcal{D}_i(w)$) of possible worlds compatible with what the agent i likes (resp. dislikes) in the world w. All these accessibility relations are serial. Moreover, for the sake of simplicity, we make the simplistic hypothesis that what is liked persists: if $w\mathcal{G}w'$ then $\mathcal{L}_i(w) = \mathcal{L}_i(w')$. Similarly for what is disliked;

– $\mathcal{I} : AGT \rightarrow (W \rightarrow 2^W)$ which associates each agent $i \in AGT$ and possible world $w \in W$ with the set $\mathcal{I}_i(w)$ of ideal worlds for the agent i. In these ideal worlds all the (social, legal, moral...) obligations, norms, standards... of agent i hold. All these relations are serial.[1]

We associate modal operators to these mappings: $After_\alpha\,\varphi$ reads "φ is true after every execution of action α", $Before_\alpha\,\varphi$ reads "φ is true before every execution of action α", $Bel_i\,\varphi$ reads "agent i believes that φ", $Prob_i\,\varphi$ reads "for i φ is more probable than $\neg\varphi$", $G\varphi$ reads "henceforth φ is true", $H\varphi$ reads "φ has always been true in the past", $Idl_i\,\varphi$ reads "ideally it is the case for i that φ" and $Des_i\,\varphi$ (resp. $Undes_i\,\varphi$) reads "φ is desirable (resp. undesirable) for i".

The truth conditions are standard for almost all of our operators: $w \Vdash \Box\varphi$ iff $w' \Vdash \varphi$ for every $w' \in \mathcal{R}_\Box(w)$ where $\mathcal{R}_\Box \in \mathcal{A} \cup \mathcal{B} \cup \{\mathcal{G}\} \cup \mathcal{I}$ and $\Box \in \{After_\alpha : \alpha \in ACT\} \cup \{Bel_i : i \in AGT\} \cup \{G\} \cup \{Idl_i : i \in AGT\}$ respectively. For the converse operators we have: $w \Vdash \boxminus\varphi$ iff $w' \Vdash \varphi$ for every w' such that $w \in \mathcal{R}_\boxminus(w')$ where $\mathcal{R}_\boxminus \in \mathcal{A} \cup \{\mathcal{G}\}$ and $\boxminus \in \{Before_\alpha : \alpha \in ACT\} \cup \{H\}$ respectively. Furthermore, $w \Vdash Prob_i\,\varphi$ iff there is $U \in \mathcal{P}_i(w)$ such that for every $w' \in U, w' \Vdash \varphi$.

Intuitively, φ is desirable for agent i if i likes φ and does not dislike φ, *viz.* φ is true in every world i likes and is false in at least one world i dislikes: $w \Vdash Des_i\,\varphi$ iff for every $w' \in \mathcal{L}_i(w), w' \Vdash \varphi$ and there is a world $w'' \in \mathcal{D}_i(w)$ such that $w'' \nVdash \varphi$. In a similar way: $w \Vdash Undes_i\,\varphi$ iff for every $w' \in \mathcal{D}_i(w), w' \Vdash \varphi$ and there is a world $w'' \in \mathcal{L}_i(w) : w'' \nVdash \varphi$. It follows from these constraints that φ cannot be simultaneously desirable and undesirable, and that there are φ that are neither desirable nor undesirable (*e.g.* tautologies and inconsistencies).

[1] We disregard thus conflicts between different kinds of standards.

We have the following introspection constraints: if $w \in \mathcal{B}_i(w')$ then $\mathcal{B}_i(w) = \mathcal{B}_i(w')$, $\mathcal{P}_i(w) = \mathcal{P}_i(w')$, $\mathcal{L}_i(w) = \mathcal{L}_i(w')$ and $\mathcal{D}_i(w) = \mathcal{D}_i(w')$, insuring that agents are aware of their beliefs, probabilities, desires, and "undesires". We also require that $U \subseteq \mathcal{B}_i(w)$ for every $U \in \mathcal{P}_i(w)$, ensuring that belief implies probability.

Dynamic operators. $After_\alpha$ and $Before_\alpha$ are defined in the standard tense logic K_t, *viz.* logic K with conversion axioms (see [16] for more details). For every α and φ, as $\mathcal{G} \supseteq \mathcal{A}_\alpha$, we have that $G\varphi \rightarrow After_\alpha\, \varphi$. As we suppose that time is linear, $Happens_\alpha\, \varphi \stackrel{def}{=} \neg After_\alpha\, \neg\varphi$ reads "α is about to happen, after which φ" and $Done_\alpha\, \varphi \stackrel{def}{=} \neg Before_\alpha\, \neg\varphi$ reads "α has just been done, and φ was true before".

In the following, the notation $i{:}\alpha$ reads "agent i is the author of action α".

Belief operators. Bel_i operators are defined in the standard KD45 logic that we do not develop here (see [17, 15] for more details).

Temporal operators. The logic of G and H is linear temporal logic with conversion axioms (see [16] for more details). $F\varphi \stackrel{def}{=} \neg G\neg\varphi$ reads "φ is true or will be true at some future instant". $P\varphi \stackrel{def}{=} \neg H\neg\varphi$ reads "φ is or was true".

Probability operators. The probability operators correspond to a notion of weak belief. It is based on the notion of subjective probability measure. The logic of $Prob$ is much weaker than the one of belief, in particular it is non-normal: the necessitation rule and the axiom K of belief operators do not have any counterpart in terms of $Prob$.

Belief and probability. They are related by the validity of:

$$Bel_i\, \varphi \rightarrow Prob_i\, \varphi \qquad \text{(BPR)}$$

We define an abbreviation $Expect_i\, \varphi$, reading "i believes that φ is probably true, but envisages the possibility that it could be false".

$$Expect_i\, \varphi \stackrel{def}{=} Prob_i\, \varphi \wedge \neg Bel_i\, \varphi \qquad (\text{Def}_{Expect})$$

Desirable/Undesirable operators. They represent preference in a wide sense. We here consider that an agent finds φ desirable, undesirable or φ leaves him indifferent. Due to the truth condition for Des_i , the following principles are valid:

if $\varphi \leftrightarrow \psi$ then $Des_i\, \varphi \leftrightarrow Des_i\, \psi$ (RE$_{Des}$)	$\neg Des_i\, \bot$ ($\bot_{Des_i}$)
$Des_i\, \varphi \rightarrow \neg Des_i\, \neg\varphi$ (D$_{Des}$)	$Des_i\, \varphi \rightarrow G Des_i\, \varphi$ (Pers$_{Des_i}$)
$\neg Des_i\, \top$ ($\top_{Des_i}$)	$\neg Des_i\, \varphi \rightarrow G\neg Des_i\, \varphi$ (Pers$_{\neg Des_i}$)
	$Des_i\, \varphi \rightarrow \neg Undes_i\, \varphi$ (RDU)

(Pers$_{Des_i}$) and (Pers$_{\neg Des_i}$) illustrate that what is desirable is atemporal. We also have corresponding principles for $Undes_i$. Note that desires are neither closed under implication nor under conjunction: I might desire to marry Ann and to marry Beth without desiring to be a bigamist. Finally, (RDU) expresses that something cannot be desirable and undesirable at the same time.

Obligation operator. The notion of obligation considered here is very wide: it embraces all the rules agents ideally have to respect. They can be explicit (like laws) or more or less implicit (like social or moral obligations). They are a kind of social preferences, imposed by a group to which the agent pertains, and thus differ from the agent's personal desires. The logic of Idl is the standard deontic logic KD (thus an agent's "obligations" must be consistent).

We will now formalize Ortony et al.'s emotions: we cite OCC's informal definition, give a formal definition, and illustrate it by OCC's examples.

3 Event-based emotions

The event-based branch of the OCC typology contains emotion types whose eliciting conditions depend on the evaluation of an event, with respect to the agent's goals. *Desirability* is a central intensity variable accounting for the impact that an event has on an agent's goals, *i.e.* how it helps or impedes their achievement. We formalize it through our Des and $Undes$ operators.

3.1 Well-being emotions

The emotion types in this group have eliciting conditions focused on the desirability for the self of an event. An agent feels joy (resp. distress) when he is pleased (resp. displeased) about a desirable (resp. undesirable) event.

$$Joy_i\varphi \stackrel{def}{=} Bel_i\,\varphi \wedge Des_i\,\varphi$$

$$Distress_i\varphi \stackrel{def}{=} Bel_i\,\varphi \wedge Undes_i\,\varphi$$

For example in [2, p. 88][2], when a man i hears that he inherits of a small amount of money from a remote and unknown relative k (Bel_i (*earn-money* $\wedge$ *k-died*)), he feels **joy** because he focuses on the desirable event (Des_i *earn-money*). Though, this man does not feel distress about his relative's death, because this is not undesirable for him ($\neg\, Undes_i$ *k-died*). On the contrary, a man j (p. 89) who runs out of gas on the freeway (Bel_j *out-of-gas*) feels **distress** because this is undesirable for him ($Undes_j$ *out-of-gas*).

3.2 Prospect-based emotions

The emotion types in this group have eliciting conditions focused on the desirability for self of an anticipated (uncertain) event, that is actively prospected. They use a local intensity variable, *likelihood*, accounting for the expected probability of the event to occur. We formalize this variable with the operator $Expect$.

[2] Below, the quoted pages all refer to OCC's book [2] so we just specify it once.

An agent feels hope (resp. fear) if he is "pleased (resp. displeased) about the **prospect** of a desirable (resp. undesirable) event"[3].

$$Hope_i\varphi \stackrel{def}{=} Expect_i\,\varphi \wedge Des_i\,\varphi$$

$$Fear_i\varphi \stackrel{def}{=} Expect_i\,\varphi \wedge Undes_i\,\varphi$$

The agent feels fear-confirmed (resp. satisfaction) if he is "displeased (resp. pleased) about the **confirmation** of the prospect of an undesirable (resp. desirable) event".

$$FearConfirmed_i\varphi \stackrel{def}{=} Bel_i\,PExpect_i\,\varphi \wedge Undes_i\,\varphi \wedge Bel_i\,\varphi$$

$$Satisfaction_i\varphi \stackrel{def}{=} Bel_i\,PExpect_i\,\varphi \wedge Des_i\,\varphi \wedge Bel_i\,\varphi$$

The agent feels relief (resp. disappointment) if he is "pleased (resp. displeased) about the **disconfirmation** of the prospect of an undesirable (resp. desirable) event".

$$Relief_i\varphi \stackrel{def}{=} Bel_i\,PExpect_i\,\neg\varphi \wedge Undes_i\,\neg\varphi \wedge Bel_i\,\varphi$$

$$Disappointment_i\varphi \stackrel{def}{=} Bel_i\,PExpect_i\,\neg\varphi \wedge Des_i\,\neg\varphi \wedge Bel_i\,\varphi$$

For example a woman w who applies for a job (p. 111) might feel **fear** if she expects not to be offered the job ($Expect_w$ *¬get-job*), or feel **hope** if she expects that she will be offered it ($Expect_w$ *get-job*). Then, if she hoped to get the job and finally gets it, she feels **satisfaction**; and if she does not get it, she feels **disappointment**. An employee e (p. 113) who expects to be fired ($Expect_e\,f$) will feel **fear** if it is undesirable for him ($Undes_e\,f$), but not if he already envisaged to quit this job ($\neg Undes_e\,f$). In the first case he will feel **relief** when he is not fired ($Bel_e\,\neg f$), and **fear-confirmed** when he is.

Theorem. We can prove some links between emotions: $Satisfaction_i\varphi \rightarrow Joy_i\varphi$ and $FearConfirmed_i\varphi \rightarrow Distress_i\varphi$. This is in agreement with Ortony et al.'s definitions. Though, we can notice that the disconfirmation-centered emotions (relief and disappointment) do not imply the corresponding well-being emotions (joy and sadness). This seems rather intuitive, since they typically do not characterize a desirable or undesirable situation, but the return to an indifferent situation that was expected to change and that finally did not.

3.3 Fortunes-of-others emotions

The emotion types in this group have eliciting conditions focused on the presumed desirability for another agent. They use three local intensity variables: *desirability for other*, *deservingness*, and *liking*. *Desirability for other* is the assessment of how much the event is desirable for the other one: for example we write $Bel_i\,Des_j\,\varphi$ for "agent i believes that φ is desirable for agent j". *Deservingness* represents how much agent i believes that agent j deserved what occurred

[3] Note that the object of hope is not necessarily about the future: I might ignore whether my email has been delivered to the addressee, and hope it has.

to him. It often depends on *liking*, *i.e.* i's attitude towards j. Below, to simplify, we assimilate "i believes that j deserves A" and "i desires that j believes A". We thus only consider *liking*, through non-logical global axioms. For example, when John likes Mary this means that if John believes that Mary desires to be rich, then John desires that Mary is rich, or rather: gets to know that she is rich ($Bel_{john}\, Des_{mary}\, rich \rightarrow Des_{john}\, Bel_{mary}\, rich$).

There are two good-will (or empathetic) emotions: an agent feels happy for (resp. sorry for) another agent if he is pleased (resp. displeased) about an event presumed to be desirable (resp. undesirable) for this agent.

$$HappyFor_{i,j}\varphi \stackrel{def}{=} Bel_i\, \varphi \wedge Bel_i\, FBel_j\, \varphi \wedge Bel_i\, Des_j\, \varphi \wedge Des_i\, Bel_j\, \varphi$$

$$SorryFor_{i,j}\varphi \stackrel{def}{=} Bel_i\, \varphi \wedge Bel_i\, FBel_j\, \varphi \wedge Bel_i\, Undes_j\, \varphi \wedge Undes_i\, Bel_j\, \varphi$$

There are two ill-will emotions: an agent feels resentment (resp. gloating) towards another agent if he is displeased (resp. pleased) about an event presumed to be desirable (resp. undesirable) for this agent.

$$Resentment_{i,j}\varphi \stackrel{def}{=} Bel_i\, \varphi \wedge Bel_i\, FBel_j\, \varphi \wedge Bel_i\, Des_j\, \varphi \wedge Undes_i\, Bel_j\, \varphi$$

$$Gloating_{i,j}\varphi \stackrel{def}{=} Bel_i\, \varphi \wedge Bel_i\, FBel_j\, \varphi \wedge Bel_i\, Undes_j\, \varphi \wedge Des_i\, Bel_j\, \varphi$$

For example (p. 95) Fred feels **happy for** Mary when she wins a thousand dollars, because he has an interest in the happiness and well-being of his friends (global axiom: $Bel_f\, Des_m\, w \rightarrow Des_f\, Bel_m\, w$). A man i (p. 95) can feel **sorry for** the victims v of a natural disaster ($Bel_i\, Bel_v\, disaster \wedge Bel_i\, Undes_v\, disaster$) without even knowing them, because he has an interest that people do not suffer undeservedly ($Undes_i\, Bel_v\, disaster$). An employee e (p. 99) can feel **resentment** towards a colleague c who receives a large pay raise ($Bel_e\, pr$, $Bel_e\, Des_c\, pr$) because he thinks this colleague is incompetent and thus does not deserve this raise ($Undes_e\, Bel_c\, pr$). Finally, Nixon's political opponents (p. 104) might have felt **gloating** about his departure from office ($Bel_o\, Bel_{nixon}\, d$, $Bel_o\, Undes_{nixon}\, d$) because they thought it was deserved ($Des_o\, Bel_{nixon}\, d$).

4 Agent-based emotions

The agent-based branch of the OCC typology contains emotion types whose eliciting conditions depend on the judgement of the praiseworthiness of an action, with respect to standards. An action is *praiseworthy* (resp. *blameworthy*) when it upholds (resp. violates) standards. We represent standards through the deontic operator Idl .

4.1 Attribution emotions

The emotion types in this group have eliciting conditions focused on the approving of an agent's action. They use two local intensity variables: *strength*

of unit[4] and *expectation deviation*. *Expectation deviation* accounts for the degree to which the performed action differs from what is usually expected from the agent, according to his social role or category[5]. We express this with the formula $\neg Prob_i\ Happens_{j:\alpha}\ \top$, reading "$i$ does not believe that it is likely that j performs successfully action α". Then $Done_{i:\alpha}\ \neg Prob_i\ Happens_{i:\alpha}\ \top$ expresses that surprisingly for himself, i succeeded in executing α. In the sequel, $Emotion_i(i{:}\alpha)$ abbreviates $Emotion_i\, Done_{i:\alpha}\ \top$ where $Emotion$ is the name of an emotion.

Self-agent emotions: an agent feels pride (resp. shame) if he is approving (resp. disapproving) of his own praiseworthy (resp. blameworthy) action.

$$Pride_i(i{:}\alpha) \stackrel{def}{=} Bel_i\ Done_{i:\alpha}\ (\neg Prob_i\ Happens_{i:\alpha}\ \top \wedge Bel_i\ Idl_i\ Happens_{i:\alpha}\ \top)$$

$$Shame_i(i{:}\alpha) \stackrel{def}{=} Bel_i\ Done_{i:\alpha}\ (\neg Prob_i\ Happens_{i:\alpha}\ \top \wedge Bel_i\ Idl_i\ \neg Happens_{i:\alpha}\ \top)$$

Emotions involving another agent: an agent feels admiration (resp. reproach) towards another agent if he is approving (resp. disapproving) of this agent's praiseworthy (resp. blameworthy) action.

$$Admiration_{i,j}(j{:}\alpha) \stackrel{def}{=} Bel_i\ Done_{j:\alpha}\ (\neg Prob_i\ Happens_{j:\alpha}\ \top \wedge Bel_i\ Idl_j\ Happens_{j:\alpha}\ \top)$$

$$Reproach_{i,j}(j{:}\alpha) \stackrel{def}{=} Bel_i\ Done_{j:\alpha}\ (\neg Prob_i\ Happens_{j:\alpha}\ \top \wedge Bel_i\ Idl_j\ \neg Happens_{j:\alpha}\ \top)$$

For example, a woman m feels **pride** (p. 137) of having saved the life of a drowning child ($Bel_m\ Done_{m:\alpha}\ \top$, where α is the action to save the child) because she thinks that her action is praiseworthy, *i.e.* its successful execution was not expected (before α, it held that $\neg Prob_m\ Happens_{m:\alpha}\ \top$) but ideally she had to perform it ($Bel_m\ Idl_m\ Happens_{m:\alpha}\ \top$). A rich elegant lady l (p. 142) would feel **shame** if caught while stealing clothes in an exclusive boutique ($Bel_l\ Done_{l:\beta}\ \top$, where β is the action to steal), because this violates a standard ($Idl_l\ \neg Happens_{l:\beta}\ \top$) and this was not expected of herself ($\neg Prob_l\ Happens_{l:\beta}\ \top$). A physicist p's colleagues c (p. 145) feel **admiration** towards him for his Nobel-prize-winning work ($Bel_c\ Done_{p:\gamma}\ \top$, where γ is the action to make some Nobel-prize-winning findings) because this is praiseworthy, *i.e.* ideal ($Bel_c\ Idl_p\ Happens_{p:\alpha}\ \top$) but very unexpected ($\neg Prob_c\ Happens_{p:\gamma}\ \top$). A man i may feel **reproach** towards a driver j (p. 145) who drives without a valid license ($Bel_i\ Done_{j:\delta}\ \top$, where δ is the action to drive without a valid license), because it is forbidden ($Bel_i\ Idl_j\ \neg Happens_{j:\delta}\ \top$), and it is not expected from a driver ($\neg Prob_i\ Happens_{j:\delta}\ \top$).

Theorem. We can prove that $Admiration_{i,i}(\varphi) \leftrightarrow Pride_i(\varphi)$ and $Reproach_{i,i}(\varphi) \leftrightarrow Shame_i(\varphi)$. This is rather intuitive, all the more Ortony et al. introduce the term *self-reproach* for shame.

[4] *Strength of unit* intervenes in self-agent emotions to represent the degree to which the agent identifies himself with the author of the action, allowing him to feel pride or shame when he is not directly the actor. For example one can be proud of his son succeeding in a difficult examination, or of his rugby team winning the championship. In this paper we only focus on emotions felt by the agent about his own actions, thus we do not represent this variable.

[5] In self-agent emotions the agent refers to the stereotyped representation he has of himself.

4.2 Composed emotions

These emotions occur when the agent focuses on both the consequences[6] of the event and its agency. They are thus the result of a combination of well-being emotions and attribution emotions.

$$\mathit{Gratification}_i(i{:}\alpha, \varphi) \stackrel{def}{=} \mathit{Pride}_i(i{:}\alpha) \wedge \mathit{Bel}_i\, \mathit{Before}_{i:\alpha}\, \neg \mathit{Bel}_i\, F\varphi \wedge \mathit{Joy}_i\varphi$$

$$\mathit{Remorse}_i(i{:}\alpha, \varphi) \stackrel{def}{=} \mathit{Shame}_i(i{:}\alpha) \wedge \mathit{Bel}_i\, \mathit{Before}_{i:\alpha}\, \neg \mathit{Bel}_i\, F\varphi \wedge \mathit{Distress}_i\varphi$$

$$\mathit{Gratitude}_{i,j}(j{:}\alpha, \varphi) \stackrel{def}{=} \mathit{Admiration}_{i,j}(j{:}\alpha) \wedge \mathit{Bel}_i\, \mathit{Before}_{j:\alpha}\, \neg \mathit{Bel}_i\, F\varphi \wedge \mathit{Joy}_i\varphi$$

$$\mathit{Anger}_{i,j}(j{:}\alpha, \varphi) \stackrel{def}{=} \mathit{Reproach}_{i,j}(j{:}\alpha) \wedge \mathit{Bel}_i\, \mathit{Before}_{j:\alpha}\, \neg \mathit{Bel}_i\, F\varphi \wedge \mathit{Distress}_i\varphi$$

For example, a woman i may feel **gratitude** (p. 148) towards the stranger j who saved her child from drowning ($\mathit{Bel}_i\, \mathit{Done}_{j:\alpha}\, \top$, where α is the action to save her child). Indeed, she feels admiration towards j because of j's praiseworthy action (*i.e.* ideal: $\mathit{Bel}_i\, \mathit{Idl}_j\, \mathit{Happens}_{j:\alpha}\, \top$, but unlikely: $\neg \mathit{Prob}_i\, \mathit{Happens}_{j:\alpha}\, \top$, for example because it needs a lot of courage). Moreover the effect of j's action (Bel_i *son-alive*) is desirable for her (Des_i *son-alive*), so she also feels joy about it (Joy_i*son-alive*). Similarly, a woman w (p. 148) may feel **anger** towards her husband h who forgets to buy the groceries ($\mathit{Bel}_w\, \mathit{Done}_{h:\beta}\, \top$, where β is his action to come back without groceries), because the result of this action ($\mathit{Bel}_w\, \neg g$) is undesirable for her ($\mathit{Undes}_w\, \neg g$), and the action was blameworthy ($\neg \mathit{Prob}_w\, \mathit{Happens}_{h:\beta}\, \top \wedge \mathit{Bel}_w\, \mathit{Idl}_h\, \mathit{Happens}_{h:\beta}\, \top$). The physicist p may feel **gratification** about winning the Nobel prize because his action γ was praiseworthy, and its result (Bel_p *is-nobel* would have been false if p had not performed γ) is desirable for him (Des_p *is-nobel*). Finally, a spy may feel **remorse** (p. 148) about having betrayed his country (action ω) if he moreover caused undesirable damages ($\mathit{Shame}_{spy}(\omega) \wedge \mathit{Distress}_{spy}\mathit{damages} \wedge \mathit{Bel}_{spy}\, \mathit{Before}_{spy:\omega}\, \neg \mathit{Bel}_{spy}\, F\mathit{damages}$.

It follows from our logic, in particular from the introspection axioms for all operators, that $\mathit{Emotion}_i\varphi \leftrightarrow \mathit{Bel}_i\, \mathit{Emotion}_i\varphi$ and $\neg \mathit{Emotion}_i\varphi \leftrightarrow \mathit{Bel}_i\, \neg \mathit{Emotion}_i\varphi$ are valid.

5 Conclusion

We have formalized twenty emotions from Ortony et al.'s theory (all but the object-based branch), thus providing a very complete set of emotions. Moreover we have shown the soundness of our framework by illustrating each definition by an example from their book. We have privileged richness, genericity, and fidelity to the definitions over tractability. An optimization would have

[6] Here, we represent the effects of an action α with the formula $\mathit{Bel}_i\, \mathit{Before}_{i:\alpha}\, \neg \mathit{Bel}_i\, F\varphi$ reading approximately "i believes that φ would not have been true if he had not performed α".

needed important concessions. For example [18] proposes a numerical model of emotions in combat games, efficient in big real-time multi-agent systems, but domain-dependant.

We would like to highlight here some shortcomings of our model. Mainly, our emotions are not quantitative (they have no intensity). This prevents us from fine-grained differentiations among emotions of the same type (for example: irritation, anger, rage). A second (and linked) shortcoming is that our emotions are persistent as long as their conditions stay true. Thereby some emotions (like Joy or $Satisfaction$) can persist *ad vitam eternam*, which is not intuitive at all. Indeed it is psychologically grounded that after an emotion is triggered, its intensity decreases, and when it is under a threshold, the emotion disappears. Finally, we cannot manage emotional blending of several emotions that are simultaneously triggered; [19] proposes an original solution to this issue. On our part, we leave these problems for further work.

References

1. Meyer, J.J.: Reasoning about emotional agents. In Proceedings of ECAI'04 (2004) 129–133
2. Ortony, A., Clore, G., Collins, A.: The cognitive structure of emotions. CUP (1988)
3. Herzig, A., Longin, D.: C&L intention revisited. In Proceedings of KR'04 (2004) 527–535
4. Herzig, A.: Modal probability, belief, and actions. Fundamenta Informaticæ **57**(2-4) (2003) 323–344
5. Darwin, C.R.: The expression of emotions in man and animals. Murray, London (1872)
6. Ekman, P.: An argument for basic emotions. Cognition and Emotion **6** (1992) 169–200
7. Russell, J.A.: How shall an emotion be called? In Plutchik, R., Conte, H., eds.: Circumplex models of personality and emotions. APA, Washington, DC (1997) 205–220
8. Becker, C., Kopp, S., Wachsmuth, I.: Simulating the emotion dynamics of a multimodal conversational agent. In: ADS'04, Springer LNCS (2004)
9. Lazarus, R.S.: Emotion and Adaptation. Oxford University Press (1991)
10. Frijda, N.H.: The emotions. Cambridge University Press, Cambridge, UK (1986)
11. Gratch, J., Marsella, S.: A domain-independent framework for modeling emotion. Journal of Cognitive Systems Research **5**(4) (2004) 269–306
12. Staller, A., Petta, P.: Introducing emotions into the computational study of social norms: a first evaluation. Journal of artificial societies and social simulation **4**(1) (2001)
13. Jaques, P.A., Vicari, R.M., Pesty, S., Bonneville, J.F.: Applying affective tactics for a better learning. In Proceedings of ECAI'04, IOS Press (2004)
14. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial Intelligence Journal **42**(2–3) (1990) 213–261
15. Chellas, B.F.: Modal Logic: an introduction. Cambridge University Press (1980)
16. Burgess, J.P.: Basic tense logic. In Gabbay, D., Guenthner, F., eds.: Handbook of Philosophical Logic. Volume VII. Second edn. Kluwer Academic Publishers (2002)
17. Hintikka, J.: Knowledge and Belief: An Introduction to the Logic of the Two Notions. Cornell University Press, Ithaca (1962)
18. Parunak, H., Bisson, R., Brueckner, S., Matthews, R., Sauter, J.: A model of emotions for situated agents. In Stone, P., Weiss, G., eds.: AAMAS'06, ACM Press (2006) 993–995
19. Gershenson, C.: Modelling emotions with multidimensional logic. In: NAFIPS'99. (1999)