A Probabilistic Formalization of the Appraisal for the OCC Event-Based Emotions

João Gluz   JCGLUZ@UNISINOS.BR

Patricia A. Jaques   PJAQUES@UNISINOS.BR

PIPCA - University of Vale do Rio dos Sinos (UNISINOS)
Av. Unisinos, 950 Bairro Cristo Rei
CEP 93.022-000 São Leopoldo, Brazil

Abstract

This article presents a logical formalization of the emotional appraisal theory, *i.e.*, it formalizes the cognitive process of evaluation that elicits an emotion. This formalization is psychologically grounded on the OCC cognitive model of emotions. More specifically, we are interested in event-based emotions, *i.e.*, emotions that are elicited by the evaluation of the consequences of an event that either happened or will happen. The formal modelling presented here is based on the AfPL Probabilistic Logic, a BDI-like probabilistic modal logic, which allows our model to verify whether the variables that determine the elicitation of emotions achieved the necessary threshold or not. The proposed logical formalization aims at addressing how the emotions are elicited by the agent cognitive mental states (desires, beliefs and intentions), and how to represent the intensity of the emotions. These are important initial points in the investigation of the dynamic interaction among emotions and other mental states.

1. Introduction

The BDI (Belief-Desire-Intention) model is a well known reasoning model in artificial intelligence research. It provides a high-level abstraction of human reasoning, allowing reasoning process modeling using only three mental states: belief, desire and intention, which represent, respectively, the informational, motivational and deliberative states of an intelligent agent (Rao & Georgeff, 1995).

The BDI is an interesting approach to model agents whose behavior relies largely on their (unobservable) mental states (Sindlar, Dastani, & Meyer, 2009), such as agents intended to display social behavior. Examples of those agents are embodied conversational agents, which are personified characters that interact with users in natural language (Cassell, 2000), and agents representing non-player characters in popular role-playing games (Sindlar et al., 2009). A socially aware agent should react in a believable way, and to do so it should take into account the presumed beliefs and goals of the agents with which it is interacting (including humans agents). Furthermore, when interacting with humans, these agents should also consider the users' emotions and their own emotions.

Endowing agents with affective reasoning allows them to be emotionally believable, *i.e.*, to show affective behavior that is consistent with the context. In addition, affective reasoning allows agents to also show social behavior. For example, it is not believable that a pedagogical agent shows to be happy for the student when the student is struggling with new content. To model social and emotional agents that show believable behavior, the

BDI model should be extended to also incorporate emotional mental states. However, most existing BDI models do not take into account the agent's emotions in its reasoning process.

At least four research questions should be considered when modeling emotional mental states in BDI agents. The first significant issue is (i) how to represent emotions and its properties, such as intensity, and (ii) how the emotions are triggered by the agent's cognitive mental states (desires, beliefs and intentions). Furthermore, as the BDI is a practical reasoning architecture, knowing (iii) how emotions can change or improve the way an agent reasons, decides or acts is also important. Another relevant aspect is (iv) how the affective mental states (*i.e.*, emotions) will influence on the cognitive ones (beliefs, desires and intentions).

This paper aims at addressing the first two issues of the list above: (i) how to represent and formalize emotion and its properties; and (ii) how emotions are elicited by the agent's cognitive mental states (desires, beliefs and intentions). Although we consider the study of the other relations between cognitive and affective mental states to be important, we chose these first two issues as an initial point of investigation, given the considerable complexity of the subject. The elicitation of emotions in particular is a promising starting point, since it allows us to explain how an emotion is triggered and it also differentiates emotions (Scherer, 2000).

We ground our emotional extension of the BDI model in the OCC model. The OCC model is a psychological appraisal model of emotion (Ortony, Clore, & Collins, 1990). It considers that an emotion is the result of an evaluation process called appraisal. For example, people may experience joy or distress when they evaluate the consequences of an event according to their goals. When the consequences of the event promote their goals, they will probably feel joy, whereas when the consequences of the event prevent them from achieving their goals, they may experience distress.

The eliciting conditions are a qualitative aspect of the OCC model, whereas the quantitative aspect of the model is represented by the intensity of the emotions (Steunebrink, Dastani, & Meyer, 2008). Emotional intensity plays an important role in the OCC model, because it allows us to differentiate experienced emotions from affective reactions. When the triggering conditions for an emotion are satisfied, the emotion has the potential to actually occur. This potential is a value that can be computed by considering the values and weights of variables that determine its intensity. An emotion will only be experienced if its potential exceeds a given threshold. When this difference is positive, it defines the intensity of the emotion. In this way, when formalizing emotions, we should also take into account their intensity.

This article presents a logical formalization of emotional appraisal, *i.e.*, it formalizes the cognitive process of evaluation that elicits an emotion according to the OCC cognitive psychological model of emotions. As stated by Ortony et al. (1990), emotions depend on the goals and preferences of the agent displaying them. Accordingly, this paper intends to present a formal model linking emotions to goals and desires in BDI agents. Moreover, due to the important role of intensity in emotions, we created a probabilistic logic, called AfPL, to formalize emotions and their intensity in our model. Although there are other works that have tried to formalize emotional BDI agents (Jiang, Vidal, & Huhns, 2007; Adam, Herzig, & Longin, 2009), they have not focused on the formalization of the intensity of emotions. The first attempt to differentiate experienced emotions from triggered emotions

by representing emotional intensity was made by Steunebrink, Dastani, and Meyer (2010, 2011). However, although they carried out a comprehensive modeling work that rigorously followed the OCC model, their method for computing intensity of emotions is limited, since it only uses BDI concepts in the calculation.

For the sake of brevity and clarity, this paper focuses on modeling event-based emotions, *i.e.*, emotions that are elicited by the evaluation of the consequences of an event for the accomplishment of a person's goals. More specifically, we are interested in the emotions that are elicited from the evaluation of the consequences for self, and we do not formalize emotions whose appraisal evaluates the consequences for others, such as resentment, pity, gloating and "happy-for". We chose to formalize the event-based emotions with consequences for self because these emotions seem to be the most important in the decision making process (Bagozzi, Dholakia, & Basuroy, 2003; Isen & Patrick, 1983; Raghunathan & Pham, 1999). Furthermore, they are also simpler to formalize, since they require less variables.

This paper is organized as follows. In Section 2, we present the OCC model, the psychological cognitive model that bases our logical formalization of the appraisal of emotions. Section 3 describes the BDI model and Section 4 presents related work and the main contributions our work brings to existing research. Section 5 explains AfPL, which is the language we used for the logical formalization of emotions. In Section 6, we describe the formalization proposed in our model for the main OCC variables that determine the emotions' intensity. Section 7 presents the proposed formalization, and Section 8 exemplifies our model with a scenario. Finally, Section 9 concludes the article, introducing ideas for future work.

2. The OCC Model

There are several psychological approaches to the study of emotions (Scherer, 2000). The dimensional models state that emotions are composed of dimensions that allow the classification of emotions, for instance according to valence and arousal. On the other hand, theorists who follow the lexical approach believe language is the main point of investigation to discover the structure of psychological phenomena. However, the cognitive view of emotion has been the most accepted approach among researchers, including in the affective computing area (Scherer, 1999; Picard, 2000; Moors, Ellsworth, Scherer, & Frijda, 2013).

According to the cognitive approach, emotions appear as a result of an evaluation process called appraisal. The central idea of the appraisal theory is that "the emotions are elicited and differentiated on the basis of a person's subjective evaluation (or appraisal) of the personal significance of a situation, event or object on a number of dimensions or criteria" (Scherer, 1999). Hence, emotions require cognitive processes to generate or to retrieve preferences or meaning.

Ortony, Clore and Collins (1990) constructed a cognitive model of emotion, known as OCC, which explains the eliciting conditions of 22 emotions types. The OCC model is an appraisal/arousal model, since it describes the eliciting conditions of emotions (appraisal) and the variables that determine their intensity (arousal). The conditions can be seen as qualitative aspects of emotions and the variables can be seen as quantitative aspects of emotions, as suggested by Steunebrink et al. (2008). For example, *hope* is qualitatively defined as "pleased about the prospect of a desirable event" (hope eliciting conditions).

The intensity of this emotion, a quantitative aspect, depends on (i) the degree to which the event is desirable, *i.e.*, its desirability, and (ii) the likelihood of the event.

2.1 The Cognitive Structure of the OCC Model

According to the OCC model, people may appraise three different aspects of the world: consequences of events, actions of agents, and aspects of objects. **Events** are the way people perceive things that happen. **Agents** can be people, biological animals, inanimate objects or abstractions such as institutions. **Objects** are objects viewed qua objects. Emotions result from *focusing on* one of these three aspects. A person can *focus on* different aspects of a same situation. For example, in the scenario given by Ortony et al. (1990), a person learns that his neighbor is a merciless child-beater. This person might only focus on the neighbor's role as the agent of child-beating (*i.e.*, on the *agent's actions*), or focus on one or more aspects of child-beating (*i.e.*, on the *event*). This person can still focus on his neighbor *qua* (unappealing) *object*. In reality, according to the OCC authors, people are likely to experience a mixture of emotions resulting from considering a situation under these different perspectives at different moments.

There are three criteria for evaluating the goodness and badness of the consequences of events, actions of agents, and aspects of objects: goals, standards, and attitudes, respectively. The consequences of events are evaluated in terms of their desirability, *i.e.*, whether they promote or thwart one's *goals* and preferences. *Standards* are used to evaluate the actions of an agent according to their obedience to social, moral, and behavioural standards or norms. Finally, the appearance of objects is evaluated in terms of appealingness according to the compatibility of their attributes with one's tastes and *attitudes*.

The elicitation of an emotion depends on a person's perception of the world – his construal. If an emotion such as distress is a reaction to an event, it is because this event is construed as undesirable. For example, when one is experiencing fear that it will rain, this is because this person construed rain as an undesirable event[1]. The event rain does not have valence, but its consequences do. This person is probably predicting that he will get wet, because he forgot his umbrella home. On the other hand, a small farmer who is loosing a harvest due to severe drought will probably construe the same rain event as a "desirable event".

Emotions are *valenced reactions*, and "any particular valenced reaction is always a reaction to one of these perspectives on the world" (*i.e.*, consequences of an event, the appearance of an object or the actions of a person) (Ortony et al., 1990). When people appraise these perspectives, they have a valenced (positive or negative) reaction. They may be *pleased* (positive) or *displeased* (negative) by the consequence of an event, they may *approve* or *disapprove* an agent's action and they may *like* or *dislike* the aspects of an object. These are general valenced reactions. A valenced reaction may or may not be experienced as an emotion, depending on how intense the reaction is and also on the specific eliciting conditions. For this reason, we can see the OCC model as an *inheritance* structure, in which the valenced reactions such as *pleased* or *displeased* are the most general type of valenced

---

1. We use the expression "(un)desirable event" in the same sense employed by OCC, *i.e.*, an event that was evaluated by someone as having (un)desirable consequences for oneself, because these consequences promote or thwart one's goals and preferences.

reaction for consequences of events, and can have different specializations according to the specific eliciting conditions someone chooses to focus on. In the rain event, for example, although the person was experiencing *fear* about getting wet, he could remember he had left a umbrella at the office the previous month and then experience *relief*. We can see fear and relief as specializations of the valenced reactions of displeased. This view of the OCC model as an *inheritance* structure was suggested by Steunebrink and colleagues (2008, 2010) and was validated by Ortony. However, we opted for following the original approach of the OCC model in this paper.

As the OCC model is not based on a theory of language of emotions, it is not concerned about the linguistic aspects of emotions, *i.e.*, which word to use to refer to every specific affective phenomenon. In the OCC model, words do not represent specific emotions, but emotion types. "An emotion type is a distinct kind of emotion that can be realized in a variety of recognizably related forms" (Ortony et al., 1990, p. 15), and which have different intensities. For example, "fear" is an emotion type that can be manifested in varying degrees of intensity, such as "concern" (less afraid), "frightened", and "petrified" (more afraid). The goal of using emotion types is to be language-neutral, so that the theory is universal, independent of culture. "Fear" is only a *token* for the type of emotions that involve being displeased about the prospective undesirable consequences of an (unconfirmed) event. The original global structure of the OCC model is illustrated in Figure 1, whose three main branches correspond to the three aspects of the world on which a person may choose to focus, originating emotions. When one focus on the **consequences of an event**, one may be *pleased* if they are desirable or *displeased* otherwise. Which emotion type is elicited depends on whether the consequences are for other or for oneself. When the consequences are for oneself (labeled as CONSEQUENCES FOR SELF in the image), the evaluation depends on whether the outcomes are actual eliciting emotions (labeled as PROSPECTS IRRELEVANT in the image), like *joy* and *distress*, or prospective (labeled as PROSPECTS RELEVANT in the image), giving rise to *hope* and *fear*. Depending on the confirmation of the prospect, four other emotions may be elicited: *satisfaction*, *disappointment*, *fear-confirmed* and *relief*. When the consequences are for other (labeled as CONSEQUENCES FOR OTHER in the image), the outcomes are evaluated according to whether they are undesirable (labeled as UNDESIRABLE FOR OTHER in the image), such as *gloating* and *pity*, or desirable (labeled as DESIRABLE FOR OTHER in the image), such as *happy-for*, *resentment*, for other.

When **actions of agents** are evaluated according to standards, valenced reactions of *approval* or *disapproval* may be elicited. The specific emotions depend on whether the action is one's own (labeled as SELF AGENT), eliciting *pride* and *shame*; or someone else's (labeled as OTHER AGENT in the image), generating *admiration* and *reproach*.

In addition to valenced reactions to either events consequences or agent's actions, there are also compound emotions that involve a *joint focus on both* the action of an agent *and* the consequences of the resulting event. These emotions correspond to valenced reactions to consequences of events caused by actions of agents (Steunebrink et al., 2011). This is the case of emotions like *gratification*, *remorse*, *anger* and *gratitude*. For example, gratitude results from "(approving of) someone else's praiseworthy action and (being pleased about) the related desirable event".
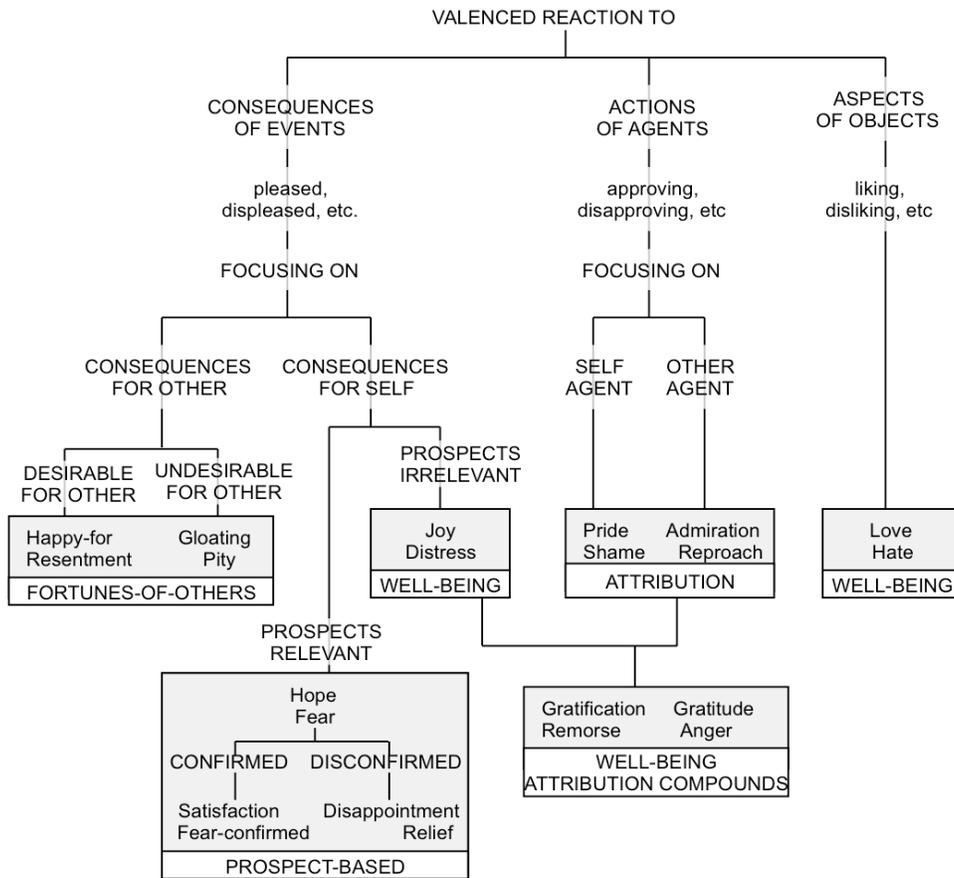
Figure 1: Global structure of emotion types - OCC Model (Ortony et al., 1990)

Finally, the **aspects of an object** are evaluated according to one's tastes and attitudes and elicit affective reactions of *liking* or *disliking*. In this case, emotions such as *love* and *hate* may be elicited. The eliciting conditions for each one of the 22 emotion types of the OCC model are described in Table 1.

As can be seen in Figure 1, the emotion types are also grouped according to their eliciting conditions. For example, the "attribution group" contains four emotion types, each of which depending on whether the attribution of responsibility for an action to an agent is positive or negative and on whether the agent is the self or another person.

The OCC authors believe that a machine that implements this model can help understand which emotions people experience and under which conditions. According to them, the objective of the OCC model is not to implement machines with emotions, but to be able to predict and explain emotion related human cognition. Conversely, Picard (2000) believes that the OCC model can be used for emotion synthesis in machines. In fact, there are already several works in computer science research that use the OCC model to implement emotions in machines (Gebhard, 2005; Dias & Paiva, 2013; Jaques, Vicari, Pesty, & Martin, 2011; Signoretti, Feitosa, Campos, Canuto, Xavier-Junior, & Fialho, 2011).

2.2 The Variables that Determine the Intensity of Event-Based Emotions

In Section 2, we describe the eliciting conditions of the 22 emotion types of the OCC model. In this section, we aim at explaining the variables that determine the intensity of an emotion. More specifically, we are going to focus only on the variables that determine the intensity of event-based emotions, which are the emotions modeled in our work.

As previously explained, the OCC model is an arousal/appraisal theory. It considers two dimensions of the emotions: the appraisal (the eliciting conditions of an emotion) and the arousal (the intensity of the emotion). Not all affective reactions (*i.e.*, evaluations of the consequences of an event as desirable or not) are necessarily emotions: "Whether or not these affective reactions are experienced as emotions depends upon how intense they are" (Ortony et al., 1990, p. 20). When the eliciting conditions of an emotion are satisfied, an emotion has the *potential* to be elicited. In fact, the intensity variables determine the potential of the emotion. The OCC assumes that there is a context-sensitive, emotion-specific *threshold* associated with each emotion. Thus, an affective reaction will be experienced as an emotion only if its potential exceeds this threshold[2]. The intensity of an emotion is given by the difference between the potential of the emotion and the threshold, when this difference is positive.

There is a primary variable that affects the intensity of all event-based emotions: **desirability**. The desirability of an event is computed according to a person's goals. If the consequences of an event promotes one's goals, the event is considered desirable, otherwise it is deemed undesirable. The degree of desirability of an event depends directly on how important the goals are for a person and on the degree to which the event in question ap-

---

2. In this article, we opted for using the expression "affective reaction" to refer to those affective reactions whose potential has not achieved a threshold to be experienced as an emotion, as used in the OCC model. Although Steunebrink (2010) refers to them as triggered emotions, we prefer to reserve the term emotion exclusively for affective reactions whose potential achieved the necessary threshold to be experienced as an emotion. Steunebrink named the latter ones "experienced emotions".

Table 1: The OCC emotion types specifications. (Adapted from Ortony et al., 1990)

Emotion	Specification
joy	pleased about the desirable consequences of an actual event
distress	displeased about the undesirable consequences of an actual event
happy-for	pleased about the consequences of an actual event presumed to have desirable consequences for someone else who one cares about
sorry-for	displeased about the consequences of an actual event presumed to have undesirable consequences for someone else who one cares about
resentment	displeased about the consequences of an actual event presumed to have desirable consequences for someone else who one does not like
gloating	pleased about the consequences of an actual event presumed to have undesirable consequences for someone else who one does not like
hope	pleased about the prospective desirable consequences of an unconfirmed event
fear	displeased about the prospective undesirable consequences of an unconfirmed event
satisfaction	pleased about the confirmation of an event with prospective desirable consequences
fears-confirmed	displeased about the confirmation of an event with prospective undesirable consequences
relief	pleased about the disconfirmation of an event with prospective undesirable consequences
disappointment	displeased about the disconfirmation of an event with prospective desirable consequences
pride	approving of one's own praiseworthy action
self-reproach	disapproving of one's own blameworthy action
appreciation	approving of someone else's praiseworthy action
reproach	disapproving of someone else's blameworthy action
gratitude	approving of someone else's praiseworthy action *and* being pleased about the desirable consequences of the resulting event
anger	disapproving of someone else's blameworthy action *and* being displeased about the undesirable consequences of the resulting event
gratification	approving of one's own praiseworthy action *and* being pleased about the desirable consequences of the resulting event
remorse	disapproving of one's own blameworthy action *and* being displeased about the undesirable consequences of the resulting event

pears to have beneficial (positive, desirable) or harmful (negative, undesirable) consequences (Ortony et al., 1990, pp. 48-49).

The intensity of positive event-based emotions increases as the positive component of the desirability increases, and the intensity of negative event-based emotions increases as the value of the negative component of the desirability variable increases.

The values of these components are generally different due to the fact that repercussions of failures and success are different. For instance, if the degree of desirability of an event is 0.7, it does not mean that its degree of undesirability is 0.3. In this way, we opted for representing these two aspects as two different variables: the degree of desirability of an event and degree of undesirability of an event.

In addition to desirability, there are other variables that determine the intensity of event-based emotions; they are: likelihood, effort, realization, desirability for other, deservingness, and liking. **Likelihood** represents the person's confidence that the event will happen. It mainly determines the intensity of prospect-based emotions, such as hope and fear. The **effort** invested by a person also influences the intensity of an emotion. A student who studied a lot for a test will be more disappointed with a low score than a student who made little effort. Another variable that determines the intensity of prospect-based emotions is the degree of **realization** of an event. For instance, still considering the school scenario, a student who had expected to score 100% in a test and scored 80% will be less frustrated than a student who had expected to score 100% and when received the test back realized that in fact he scored 30%.

The three last local variables determine the intensity of emotions from the "fortune-of-others" group. The emotions in this group (*e.g.*, resentment, gloating, happy-for, pity) are affective reactions to the consequences of desirable or undesirable events for another person. The degree to which the consequences of an event are seen as **(un)desirable for the other person** also determines the intensity of these emotions. The degree of **liking** one has for the other person also contributes to the intensity of fortune-of-others emotions. For instance, one will more likely be happy about the promotion of a colleague if they are friends. Lastly, the intensity of fortune-of-others emotions are also influenced by one's perceptions of the degree to which the event affecting the other person is fair or deserved, *i.e.*, the person's **deservingness**. Like desirability-for-other, deservingness is estimated from the point of view of the person experiencing the emotion. For example, one may experience pity for a coworker who was fired, because one evaluates the consequences of the event as undesirable for the coworker. The coworker, however, might have wanted to be fired for personal reasons and is actually happy to be sent home.

There are also global variables that determine the intensity of all emotions; they are: sense of reality, proximity and unexpectedness. **Sense of reality** relates to the degree to which the event seems real to the person experiencing the emotion (Ortony et al., 1990, p. 60). Simply imagining an event is not sufficient to elicit an emotion. For example, imagining winning a fortune in a lottery is not likely to elicit happiness. The second variable represents the psychological **proximity** of the event. Situations tend to generate more intense emotions when, for instance, they are close in time or space. This variable determines the intensity of emotions that are elicited by reviving memories from the past as well as by imagining future situations. In the first case, the intensity of the corresponding emotions diminishes as the past situation recedes into the past. The second case determines the intensity of prospect-based emotions. The closer in time the prospect is to happen, the more intense the emotion. Proximity refers not only to closeness in time, but also to other types of psychological proximity, such as spatial proximity. For example, people living close to the site of a nuclear power plant accident would probably be more afraid than other people living farther away (Ortony et al., 1990, p. 116). The third variable is **unexpected** and it

represents how unlikely it is that an event would happen. Unexpected events generate more intense emotions than expected events. For instance, consider the case of a student who for the first time scored 100 out of 100; this student will be happier than another student who usually gets 100. The OCC model also presents the **arousal** variable, however, as it is a physiological variable and not a cognitive one, it will not be taken into consideration in this work.

As aforementioned, the intensity variables explained in this chapter determine the potential of an emotion. An emotion will be experienced by a person only if its potential exceeds the emotion-specific threshold. The *intensity* of an emotion is defined as the difference between the magnitude of its potential and the current value of the emotion threshold (Ortony et al., 1990, p. 82). The OCC authors suggest that this potential is calculated by considering the value and the weight of each variable that affects emotion intensity. The *value* of the variable can be scaled from -1 to +1. For example, in the variable liking, -1 represents the maximum disliking, +1 the maximum liking, and 0 represents neither liking or disliking the person in question. About the variable *weight*, the authors claim that it can vary from an emotion type to another or even from an emotion token to another within each type. However, they do not explain how to assign a variable weight for an emotion type (Ortony et al., 1990, p. 83). Therefore, we opted for not taking into account the variable weight in our formalization. Thus, all variables in our model influence the intensity of an emotion equally, i.e., all variables have the same weight. This could cause our model to overestimate or underestimate the importance of some variables when calculating the intensity of an emotion.

3. The BDI Model

The BDI (Belief-Desire-Intention) model is based on the notion of mental states to describe the behavior of a cognitive agent. This notion was initially described in the works of Dennett (1971, 1997) and Searle (1998), and it was later extended by the philosopher Michael Bratman (1990), who gave particular attention to the role of intentions in reasoning (Wooldridge, 1999).

The BDI approach models an intelligent agent as having certain mental attitudes of belief, desire, and intention, which represent, respectively, the information, motivational and deliberative states of the agent (Rao & Georgeff, 1995).

The **beliefs** represent the information about the state of the environment. Beliefs can be viewed as the informative component of the system state. For example, consider a robotic vacuum cleaner that inhabits an environment (room) composed by a grid with two cells (Russell & Norvig, 2010). There is a unique "home" cell for the robot, and the other location may be clean or may contain dirt. The agent can detect whether the home location is dirty or not and in which cell it is located. The goal of the vacuum cleaner agent is to traverse the room, clean up all the dirt, and return home. Possible beliefs for this agent are: cell A is clean; cell B is dirty; I'm located in the cell B; my battery energy level is X; etc.

The **desires** are the motivational state of the system. They have information about the objectives to be accomplished, *i.e.*, what priorities or payoffs are associated with the various current objectives. Desires represent situations that an agent wants to achieve. The fact

that an agent has a desire does not mean that the agent will satisfy it. The agent carries out a deliberative process in which the agent confronts its desires and beliefs and chooses a set of desires that can be satisfied. For example, the vacuum cleaner agent should have as desires: (i) if my battery energy level is low, I should recharge myself immediately; (ii) if I'm located in a dirty cell, I should clean the cell; (iii) if I'm located in a clean cell, I should move to the next cell, etc.

A **goal** is a desire that the agent chooses for active pursuit. Goals should be consistent (Rao & Georgeff, 1995). Desires can be contradictory to each other, but the goals cannot (van der Hoek & Wooldridge, 2012, p. 98). Thus, an agent can desire P and $\neg P$ at the same time, but only one of the two can become a goal. For instance, if the robotic vacuum cleaner has the desires of cleaning a cell and recharging itself, and the recharging implies that it cannot execute cleaning activities, then the agent cannot have both goals at the same time.

An **intention** is a goal that is chosen to be executed by a plan (Rao, 2009; Sadek, 1992). This choice is made because the agent believes it can satisfy the goal (it is not rational for an agent to carry out something that it does not believe it can achieve). Plans are procedures that depend on a set of conditions for being applicable (Rao, 2009). Intentions are persistent and represent the currently chosen course of action. An agent will not give up on its intentions – they will persist, until the agent believes it has successfully achieved them, it believes it cannot achieve them or the purpose of the intention is no longer present. For example, the vacuum cleaner agent will choose as intention the goal to clean the cell if the cell is dirty and the cleaner has enough battery power.

A BDI agent performs actions that it intends to execute without any further reasoning, until it is forced to revise its own intentions due to changes in its beliefs or desires. This may happen because of new events or the failure or successful conclusion of existing intentions. In the example of the cleaning robot above, as soon the current cell is clean the agent will abandon the intention of cleaning this cell.

4. Related Works

Other researchers have also worked on the extension of the BDI architecture to incorporate emotions. Jiang et al. (2007) defines an extension of the generic software model of BDI agents that introduces emotions. Their work explores the possible influences of emotions in determining agents' beliefs and intentions. However, it does not have a specific formalization of the appraisal process of emotion. Van Dyke Parunak, Bisson, Brueckner, Matthews, and Sauter (2006) propose an extension of the BDI software model to integrate emotions, which is based on the OCC model. In this model, the agent's beliefs influence on its appraisal, which determines its emotions. The emotions, on the other hand, affect in the choice of intentions. Again, this work does not propose a formal representation of the model and its focus is on the impact of the emotions in the agent's choice of intentions from desires. Another work that takes a computational approach to integrate emotions into BDI agents is the EMA (Emotion and Adaptation) framework (Gratch & Marsella, 2004, 2005). The EMA framework aims at representing both the emotional appraisal and the impact emotions have on cognitive and behavioral responses, particularly coping responses. The works cited here propose extensions of the generic software model of BDI agents to integrate emotions.

However, differently from our work, they are not interested in the formalization of the appraisal process of emotions.

A more closely related work is the one proposed by Adam et al. (2009). Although the work also presents a logical formalization of the OCC model, we observe three main differences. First of all, they use a BDI modal logic to describe the OCC model, while in the proposed work we extend a BDI modal logic to incorporate emotions through a new operator of desirability. It means that Adam et al.'s work is a logical formal representation of emotions, and the proposed work aims at representing the emotional mental states in the BDI logic by formalizing their eliciting conditions. In the future, our approach will enable the study of the consequences of emotions on other mental states and its contrariwise effects. Besides, Adam and colleagues' work does not distinguish between emotion elicitation and emotion experience. Their work does not use a probabilistic BDI logic to represent the OCC emotion elicitation, thus their model is not able to represent whether the potential of an emotion in the appraisal process achieved the threshold to elicit the emotion. So, according to their model, in the case that the conditions for an emotion are met, the user will always experience it. This does not follow the OCC model, which establishes thresholds for an emotion to be elicited (Ortony et al., 1990). Furthermore, Adam et al. use desirability and desires as synonymous, which we do not believe to be the most appropriate approach. From our point of view, desires are what the OCC theory calls goals, and they have an important role, since a person uses his goals to evaluate the desirability of an event, *i.e.*, the desirability of the consequences of an event is evaluated according to one's goals. It is the desires that will determine whether the consequences of an event are desirable or not.

Steunebrink et al. (2010, 2011) have also formalized the OCC model using a BDI logic. Their work consists in the formalization of three phases of emotion: elicitation, experience and regulation. This formalization is made in three stages, such that each stage is built on the previous stage. The first stage aims at modelling the conditions that trigger emotions in a semiformal way. The second stage represents the main notions of the OCC model in a dynamic doxastic logic. The last stage uses a BDI-framework with achievement goals to represent the three phases of emotion. According to the authors, this formalization in three stages allows the system to "provide different levels of commitment to formalisms, so that it facilitates to extend or replace the used formalisms without having to start from scratch" (Steunebrink et al., 2011). Their work is a first attempt to differentiate affective reactions from experienced emotions by representing emotions' intensity. Although they represented the variables that determine the emotions' intensity and the notions of emotion threshold and potential, which are necessary to represent emotions' intensity, they did not discuss possible calculations for all variables affecting intensity.

In the proposed formalization, which is described in the next section, we intend to handle the presented drawbacks by defining the notion of desirability and other variables that determine the intensity of emotions and working with a probabilistic extension of the BDI model to calculate and represent the intensity of the event-based emotions of the OCC model. These aspects allow us to differentiate experienced emotions (emotions that are actually being experienced by a person) from affective reactions (reactions that have not achieved the threshold to be experienced as an emotion, *i.e.*, reactions that have such low potential that it did not allow them to be experienced by an individual).

5. The *AfPL* Logic

In order to formalise the OCC model, we defined a new logic, called *AfPL* (*Affective Probabilistic Logic*), that extends classical modalities of the BDI model to work within a probabilistic framework. The *AfPL* logic was carefully designed to contain the minimum set of modalities necessary to fully represent the OCC model in a BDI context. The elementary part of *AfPL* is a propositional logic with the traditional logic operators. If φ and ψ are *AfPL* formulas, then $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \to \psi$, and $\neg \varphi$ are also *AfPL* formulas. Upon this elementary part, several other linguistic constructs were added to represent the basic properties of the OCC and BDI models.

The ability of the *AfPL* to represent perceptions and actions, including sequences of actions (*i.e.*, plans), is one of the constructs abovementioned. For both the OCC and BDI models, perceptions and actions are important concepts, allowing the understanding of how emotions or rational behaviors arise. *AfPL* allows primitive *events* and *actions* to be combined in sequences through event expressions ε in the form $e_1; e_2; ...; e_n$. Formally, actions are a subset of events explicitly related to the agent. Informally, actions represent events (*i.e.*, modifications in the state of the environment) caused by the agent. This allows the separation between things that happen in the environment (*i.e.*, not caused by the agent) and things that can be caused by the agent and hence are the basis for the agent's plans. Actions are a subset of events, thus it is possible to check whether a sequence of events ε is formed only by actions using the predicate $\boldsymbol{Act}(\varepsilon)$.

A temporal modality was necessary, because there are several time-related aspects of the emotional appraisal process in the OCC model. For instance, the desirability of an event is clearly related to how much influence this event has in the future achievement of the agent's goals. To handle the time-related aspects of the OCC model, *AfPL* includes a linear temporal modality (Hodkinson & Reynolds, 2007). The linear model of time provides the necessary expressibility power to represent time concepts in the OCC model, *i.e.*, we do not need to consider multiple time lines or paths to formalize this model.

Temporal assertions about a condition φ can be made before or after an event e occur. The *before* operator $\boldsymbol{Bf}(e, \varphi)$ is true if the condition φ is hold on the previous state of the world, just before the occurrence of the event e. The *after* operator $\boldsymbol{Af}(e, \varphi)$ is true if φ is satisfied on the next state of the world, just after event e happens.

The only form of quantification allowed in *AfPL* is the quantification over event variables. This allows the definition of, for instance, the modality $\diamond \varphi$, which asserts that φ will eventually be true, by the *AfPL* expression $(\exists x)\boldsymbol{Af}(x, \varphi)$, or to express that all sequences of events have a next event: $(\forall x)(\exists y)(\boldsymbol{Bf}(x, \top) \to \boldsymbol{Af}(y, \top))$, where $\top$ is any valid formula.

In accordance to the BDI model, *AfPL* allows the agent to have *beliefs* and *goals*, which are non contradictory desires (see Section 3). Beliefs provide the basic informational component of agents' mental states. To allow the agent to express beliefs, *AfPL* supports the belief modal operator $\boldsymbol{B}(\varphi)$, which is defined as a KD45 modality (Hughes & Cresswell, 2012). Goals are the basic motivational attitudes of the agent and are expressed by the $\boldsymbol{G}(\varphi)$ modal operator, defined as a KD modality (Hughes & Cresswell, 2012). In this aspect, *AfPL* follows the same KD modality used in Standard Deontic Logics (SDL) (Meyer & Veltman, 2007) for obligatory modal operator. Operators for *persistent goals* and *intentions*,

commonly used in BDI logics, can also be defined in *AfPL* by syntactic means (see Section 5.3 for details).

BDI beliefs and goals are necessary, but they are not enough to capture all details of the OCC model. Some elementary variables of this model, like the desirability of an event, depend on the value of the goals affected by the event, and on the degree that the consequences of this event will advance the achievement of those goals[3]. In the OCC model, these values and degrees are effectively represented as real values, so, in our opinion, to give a trustworthy formal model for these variables, the underlying logic needs also to be able to represent real numbers and make inferences with them.

Probabilities, in the form of its subjective or Bayesian interpretation (see Ghosh, De-lampady, & Samanta, 2006), were also considered necessary, because they provide a credible and measurable model for the consequences of events, which is a requirement in the OCC model's estimation process of elementary variables, like desirability or likelihood. The sub-jective interpretation of probabilities is consistent with the agent making estimations about how much it believes in something. The subjective probability that the agent has in a belief φ is denoted by the term $\boldsymbol{bd}(\varphi)$. Intuitively, this term informs the *belief degree*, and can be used by the agent to make a subjective estimation about the probability of φ to be true. For example, $\boldsymbol{bd}((\exists x)\boldsymbol{Af}(x, p)) < \boldsymbol{1}$ asserts a not so firm belief that p will be true after an event occur, assigning a belief degree lesser than one for this belief.

The value assigned to a goal is used to measure how much the agent wants the outcome related to this goal, or how much this outcome will be beneficial, or desirable to it. Intu-itively, this value will inform the *force* or the *intensity* of the desire to achieve the goal. In *AfPL*, the *desire force* assigned to φ goal, will be denoted by the term $\boldsymbol{df}(\varphi)$. A desire force of zero indicates no desire at all to achieve the goal. Desire forces greater than zero indicate crescent intensities for desire to achieve the goal. To keep desire forces measurable, we assume that there are not unlimited (infinite) desire forces in *AfPL*, *i.e.*, thus there exists a maximum force to be assigned to each desire.

AfPL also provides a logical test operator $\boldsymbol{?}(\varphi)$, which tests the logical value of φ, returning 1 if φ is true, and 0 if it is false. This operator allows the definition of functions with conditional expressions directly in *AfPL* .

To work with probabilities, desire forces, and numerical real values, *AfPL* incorporates the algebraic theory of the Real Closed Field (RCF) (Monk, 1976; Warner, 1990, p. 362, p. 493). The RCF has a complete axiomatization, due to a result from Tarski (Monk, 1976, p. 362), and this axiomatization is partially incorporated into *AfPL*[4].

5.1 Syntax of AfPL

The abstract syntax of *AfPL* is defined for φ formulas, ε event expressions, and π numeric expressions as follows.

---

3. In this context, we are referring to OCC goals. The difference between OCC goals and BDI goals, as well their definition in the *AfPL* logic, is presented in Section 6.1.
4. *AfPL* cannot express polynomials or polynomial equations because it does not support numerical vari-ables and quantifiers over numerical variables. Thus, the axiomatization of the numerical component of *AfPL* cannot be complete, because the axioms that assure that all positive real numbers must have a root, and that every odd degree polynomial also has a root, cannot be expressed.

$$\varphi ::= (\varphi) \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid (\exists x)\varphi \mid \boldsymbol{B}(\varphi) \mid \boldsymbol{G}(\varphi) \mid p \mid \pi \leqslant \pi \mid$$
$$\boldsymbol{Act}(\varepsilon) \mid \boldsymbol{Af}(\varepsilon, \varphi) \mid \boldsymbol{Bf}(\varepsilon, \varphi)$$
$$\varepsilon ::= e \mid x \mid \varepsilon; \varepsilon$$
$$\pi ::= \pi + \pi \mid \pi \times \pi \mid -\pi \mid \pi^{-1} \mid \boldsymbol{0} \mid \boldsymbol{1} \mid \boldsymbol{bd}(\varphi) \mid \boldsymbol{df}(\varphi) \mid \boldsymbol{?}(\varphi)$$

In the syntax above, the symbols p, e, and x represent, respectively, the atomic propositions, atomic events, and event expression variables of *AfPL*.

Several commonly used logical, numerical, and modal operators are defined by purely syntactical means. The remaining logical operators and the universal quantification are defined as follows:

$$\varphi \rightarrow \psi \stackrel{\text{def}}{=} \neg\varphi \vee \psi \qquad \varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi) \qquad (\forall x)\varphi \stackrel{\text{def}}{=} \neg(\exists x)\neg\varphi$$

The numerical equality and the remaining comparing operators are defined as follows:

$$r = s \stackrel{\text{def}}{=} (r \leqslant s \wedge s \leqslant r) \qquad r \neq s \stackrel{\text{def}}{=} \neg(r = s)$$
$$r < s \stackrel{\text{def}}{=} (r \leqslant s \wedge r \neq s) \qquad r \geqslant s \stackrel{\text{def}}{=} s \leqslant r \qquad r > s \stackrel{\text{def}}{=} s < r$$

Numerical subtraction and division can be easily defined with the correspondent inverse operator as follows:

$$r - s \stackrel{\text{def}}{=} r + -s$$
$$r \div s \stackrel{\text{def}}{=} r \times s^{-1}$$

It is possible to work with conditional probabilities in *AfPL*. The $\boldsymbol{bd}(\varphi|\psi)$ operator is the conditional probability operator of *AfPL* equivalent to the $P(A|B)$ operator of Probability Theory. This operator is defined for subjective probabilities (*i.e.*, belief degrees) as follows:

$$\boldsymbol{bd}(\varphi \mid \psi) \stackrel{\text{def}}{=} \boldsymbol{bd}(\varphi \wedge \psi) \div \boldsymbol{bd}(\psi)$$

5.2 Semantics of *AfPL*

The model-based semantics for *AfPL* formulas is defined over models M with the following structure:

$$M = \langle \mathcal{W}, \mathcal{E}, \mathcal{A}, \mathcal{B}, \mathcal{G}, \mathcal{T}, \mathcal{I}, \mathcal{P}, \mathcal{M}, \mathcal{R} \rangle$$

The non-empty set $\mathcal{W}$ provides the possible worlds in which *AfPL* expressions will be evaluated. The non-empty sets $\mathcal{E}$ and $\mathcal{A}$ define, respectively, the set of events that can occur in the environment, and the actions the agent can perform in the environment. Actions are a subset of events, $\mathcal{A} \subseteq \mathcal{E}$, so the agent can perceive its own actions.

The relations $\mathcal{B} \subseteq \mathcal{W} \times \mathcal{W}$, and $\mathcal{G} \subseteq \mathcal{W} \times \mathcal{W}$ define the accessibility relations among the possible worlds of $\mathcal{W}$, based on the beliefs ($\mathcal{B}$) and goals ($\mathcal{G}$) of the agent. The restriction over $\mathcal{B}$ is that this relation must be serial: for every $w \in W$, there is a $v \in W$, such that $w\mathcal{B}v$ (the D modal property), transitive: for all $w, v, u \in W$, if $w\mathcal{B}v$ and $v\mathcal{B}u$ then $w\mathcal{B}u$ (the 4 modal property) and Euclidean: for all $w, v, u \in W$ if $w\mathcal{B}v$ and $w\mathcal{B}u$ then $u\mathcal{B}v$ (the 5 modal property), being, thus, an epistemic (*doxastic*) KD45 modality (Hughes & Cresswell, 2012), whereas $\mathcal{G}$ is a serial relation defining the *deontic* KD modality of the

logic (Meyer & Veltman, 2007; Hughes & Cresswell, 2012). Expressions $w\mathcal{B}v$ and $w\mathcal{G}v$ are compact notations, respectively, for $(w,v) \in \mathcal{B}$ and $(w,v) \in \mathcal{G}$. Sets $\mathcal{B}(w)$ and $\mathcal{G}(w)$ are defined as $\mathcal{B}(w) = \{w' \in \mathcal{W} | w\mathcal{B}w'\}$ and $\mathcal{G}(w) = \{w' \in \mathcal{W} | w\mathcal{G}w'\}$.

The labeled transition relation $\mathcal{T} \subseteq \mathcal{W} \times \mathcal{E} \times \mathcal{W}$ defines a linear temporal relation among possible worlds, specifying the transitions between worlds caused by events or actions (Hodkinson & Reynolds, 2007). The notation $w \xrightarrow{e} v$ stands for $(w,e,v) \in \mathcal{T}$, when $e \in \mathcal{E}$, and $w,v \in \mathcal{W}$. This notation can be extended to transition sequences, thus $w \xrightarrow{e_1} u_1 \xrightarrow{e_2} u_2...u_{n-1} \xrightarrow{e_n} v$, represents $w \xrightarrow{e_1} u_1$, $u_1 \xrightarrow{e_2} u_2$, ..., $u_{n-1} \xrightarrow{e_n} v$. The relation $\mathcal{T}$ must define a linear flow of time, thus its underlying time relation $\mathcal{T}_< = \{ (v,w) \mid v \xrightarrow{e_1} u_1 \xrightarrow{e_2} u_2...u_{n-1} \xrightarrow{e_n} w \in \mathcal{T}$ for a $e_1,e_2,...,e_n \in \mathcal{E}$, $u_1,u_2,u_{n-1} \in \mathcal{W}$, and $n>0 \}$ must be transitive and antisymmetric (Hodkinson & Reynolds, 2007).

The mapping $\mathcal{I}$ denotes a set of interpretations $\mathcal{I}_w$, possibly distinct for each world $w \in \mathcal{W}$. Each interpretation $\mathcal{I}_w$ is formed by a set of atomic propositions of *AfPL*.

The mapping $\mathcal{P}$ assigns to each world $w \in \mathcal{W}$ a probability space $\mathcal{P}_w$ with structure $\mathcal{P}_w = \langle \mathcal{W}_w, \mathcal{F}_w, \mu_w \rangle$ (for details, see Fagin & Halpern, 1994; Halpern, 2005), in which $\mathcal{W}_w \subseteq \mathcal{B}(w)$ is the subset of $\mathcal{B}(w)$ that provides the sample space of $\mathcal{P}_w$, $\mathcal{F}_w$ is a σ-algebra over the subsets of the sample space $\mathcal{W}_w$, and μ_w is a probability distribution function over this sample space that assigns for each world $w' \in \mathcal{W}_w$ a probability value $\mu_w(w')$. Each function μ_w respects the Kolmogorov Axiom of non-negativity of probabilities: $0 \leqslant \mu_w(w')$ for all $w' \in \mathcal{W}_w$, of normalization of probabilities: $\mu_w(\mathcal{W}_w) = 1$, and of enumerable additivity: $\mu_w(\mathcal{W}_w^1 \cup \mathcal{W}_w^2 \cup ...) = \mu_w(\mathcal{W}_w^1) + \mu(\mathcal{W}_w^2) + ...$, for all $\mathcal{W}_w^i \subseteq \mathcal{W}_w$, and $\mathcal{W}_w^j \subseteq \mathcal{W}$, such that $\mathcal{W}_w^i \cap \mathcal{W}_w^j = \emptyset$.

The mapping $\mathcal{M}$ assigns to each world $w \in \mathcal{W}$ a measurement space (Halmos, 1974) $\mathcal{M}_w$ with structure $\mathcal{M}_w = \langle \mathcal{U}_w, \mathcal{H}_w, \omega_w \rangle$, where $\mathcal{U}_w \subseteq \mathcal{G}(w)$ is the subset of $\mathcal{D}(w)$ that provides the measurement space of $\mathcal{M}_w$, $\mathcal{H}_w$ is a σ-algebra over the subsets of the measurement space $\mathcal{U}_w$, and ω_w is a finite measure function over this measurement space that assigns for each world $w' \in \mathcal{U}_w$ a real non-negative value $\omega_w(w')$. To be a finite measure function over the set of possible worlds $\mathcal{U}_w$, each ω_w must be non-negative: $0 \leqslant \omega_w(w')$ for all $w' \in \mathcal{D}_w$, finite: there is a real value $\hat{\omega} \in \mathcal{R}$ such that $\omega_w(\mathcal{U}_w) \leqslant \hat{\omega}$, and enumerable additive: $\omega_w(\mathcal{U}_w^1 \cup \mathcal{U}_w^2 \cup ...) = \omega(\mathcal{U}_w^1) + \omega(\mathcal{U}_w^2) + ...$, for all $\mathcal{U}_w^i \subseteq \mathcal{U}_w$, and $\mathcal{W}_w^j \subseteq \mathcal{U}_w$, such that $\mathcal{W}_w^i \cap \mathcal{U}_w^j = \emptyset$ (Halmos, 1974).

The totally ordered field of real numbers $\mathcal{R}$ provides the domain for the numerical and probabilistic expressions of *AfPL*. It is equivalent (up to isomorphism) to the Real-Closed Field (Warner, 1990, p. 493), including addition $r +_\mathcal{R} s$, multiplication $r \times_\mathcal{R} s$, negation $-_\mathcal{R} r$, reciprocal $r^{-1\mathcal{R}}$ and the ordering relation $r \leqslant_\mathcal{R} s$, which are defined over the real numbers r and s. It also includes the neutral elements of addition $(0_\mathcal{R})$ and multiplication $(1_\mathcal{R})$.

The formal semantics of *AfPL* formulas are defined through the logical consequence relation $\models$ maintained among M models, w worlds, and *AfPL* formulas. In the following definitions, φ and ψ represent well-formed formulas (*wff*); ε, ε_1, ε_2 ,... represent syntactic variables over event and action expressions; e, e_1, e_2, ..., represent primitive events, w, v, u, u_1, u_2 , ... represent possible worlds, x and p represent, respectively, event expression variables and logical propositions; π, π_1, π_2,... represent numerical (real) terms; $\varphi[x/e_1; e_2; ...; e_n]$ represents the substitution in φ *wff* of all occurrences of x variable by the $e_1; e_2; ...; e_n$ action expression, formed by the sequence of primitive events $e_1, e_2, ..., e_n$.

$$M,w \models (\varphi) \text{ iff } M,w \models \varphi$$
$$M,w \models \neg\varphi \text{ iff } M,w \not\models \varphi$$
$$M,w \models \varphi \wedge \psi \text{ iff } M,w \models \varphi \text{ and } M,w \models \psi$$
$$M,w \models \varphi \vee \psi \text{ iff } M,w \models \varphi \text{ or } M,w \models \psi$$
$$M,w \models p \text{ iff } p \in \mathcal{I}_w$$
$$M,w \models \boldsymbol{B}(\varphi) \text{ iff } M,v \models \varphi, \text{ for all worlds } v \text{ such that } w\mathcal{B}v$$
$$M,w \models \boldsymbol{G}(\varphi) \text{ iff } M,v \models \varphi, \text{ for all worlds } v \text{ such that } w\mathcal{G}v$$
$$M,w \models \exists x(\varphi) \text{ iff } M,w \models \varphi[x/e_1; e_2; ...; e_n]$$
$$\text{for some } e_1, e_2,..., e_n \in \mathcal{E} \text{ with } n > 0 \tag{1}$$
$$M,w \models \boldsymbol{Af}(e_1; e_2; ...; e_n, \varphi) \text{ iff } M,v \models \varphi$$
$$\text{for all } w \xrightarrow{e_1} u_1 \xrightarrow{e_2} u_2...u_{n-1} \xrightarrow{e_n} v \in \mathcal{T}$$
$$\text{such that } u_1, u_2, ..., u_{n-1}, v \in \mathcal{W} \text{ and } e_1, e_2,, ..., e_n \in \mathcal{E}$$
$$M,w \models \boldsymbol{Bf}(e_1; e_2; ...; e_n, \varphi) \text{ iff } M,v \models \varphi$$
$$\text{for all } v \xrightarrow{e_1} u_1 \xrightarrow{e_2} u_2...u_{n-1} \xrightarrow{e_n} w \in \mathcal{T}$$
$$\text{such that } v, u_1, u_2, ..., u_{n-1} \in \mathcal{W} \text{ and } e_1, e_2,, ..., e_n \in \mathcal{E}$$
$$M,w \models \boldsymbol{Act}(\varepsilon) \text{ iff for all events } e \text{ on } \varepsilon, e \in \mathcal{A}$$
$$M,w \models \pi_1 \leqslant \pi_2 \text{ iff } [\![\pi_1]\!]_{M,w} \leqslant_{\mathcal{R}} [\![\pi_2]\!]_{M,w}$$

This semantic needs one additional function to be fully defined: the $[\![\,]\!]$ term evaluation function, which specifies the semantic of numerical, and probabilistic terms of the $\leqslant$ ordering relation. This function is defined by the following set of equations.

$$[\![\pi_1 \times \pi_2]\!]_{M,w} = [\![\pi_1]\!]_{M,w} \times_{\mathcal{R}} [\![\pi_2]\!]_{M,w}$$
$$[\![\pi_1 + \pi_2]\!]_{M,w} = [\![\pi_1]\!]_{M,w} +_{\mathcal{R}} [\![\pi_2]\!]_{M,w}$$
$$[\![-\pi]\!]_{M,w} = -_{\mathcal{R}} ([\![\pi]\!]_{M,w})$$
$$[\![\pi^{-1}]\!]_{M,w} = ([\![\pi]\!]_{M,w})^{-1_{\mathcal{R}}}, \text{ if } [\![\pi]\!]_{M,w} \neq 0$$
$$[\![\boldsymbol{0}]\!]_{M,w} = 0_{\mathcal{R}}$$
$$[\![\boldsymbol{1}]\!]_{M,w} = 1_{\mathcal{R}} \tag{2}$$
$$[\![\boldsymbol{bd}(\varphi)]\!]_{M,w} = \sum_{v \in \mathcal{W}_w \text{ and } M,v \models \varphi} \mu_w(v)$$
$$[\![\boldsymbol{df}(\varphi)]\!]_{M,w} = \sum_{v \in \mathcal{U}_w \text{ and } M,v \models \varphi} \omega_w(v)$$
$$[\![\boldsymbol{?}(\varphi)]\!]_{M,w} = \begin{cases} 1 & \text{if } M, w \models \varphi \\ 0 & \text{otherwise} \end{cases}$$

The operator $\boldsymbol{?}(\varphi)$ maps the logical value of φ to real numbers 0 and 1. It returns 1 (true) if φ is satisfied in the current world w of model M, and returns 0 (false) otherwise. For the semantic of belief probabilities, we followed Halpern's approach on probabilistic

logics (Halpern, 1990). However, we restricted it to the assignment of probabilities of propositional formulas. Belief probabilities are defined by the application of μ, the discrete probability distribution function of the agent, over the possible worlds that can be reached through the belief accessibility relation. The semantic of desire forces is similarly defined by the application of ω to the worlds accessible by the desire's accessibility relation.

5.3 Additional Operators

Additional temporal operators can be easily defined with the help of the $\boldsymbol{Bf}(e, \varphi)$ and $\boldsymbol{Af}(e, \varphi)$ operators. The $\boldsymbol{Next}(\varepsilon)$ and $\boldsymbol{Last}(\varepsilon)$ operators, which verify if an event ε will happen next or has happened immediately before are defined as:

$$\boldsymbol{Next}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{Af}(\varepsilon, \top)$$

$$\boldsymbol{Last}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{Bf}(\varepsilon, \top)$$

The operator that tests for the previous occurrence of a condition ($\boldsymbol{Prev}(\varphi)$), and the converse operator ($\boldsymbol{Poss}(\varphi)$) for the future possibility modality ($\diamond\varphi$), are defined as follows:

$$\boldsymbol{Prev}(\varphi) \stackrel{\text{def}}{=} (\exists x)\boldsymbol{Bf}(x, \varphi)$$

$$\boldsymbol{Poss}(\varphi) \stackrel{\text{def}}{=} (\exists x)\boldsymbol{Af}(x, \varphi)$$

The $\boldsymbol{Later}(\varphi)$ operator asserts that although some condition φ is not true now, it eventually will be true, and the $\boldsymbol{Before}(\varphi, \psi)$ operator states that condition φ is true before condition ψ. Both operators were defined according to, respectively, ($LATER\ x\ p$) and ($BEFORE\ x\ p\ q$) operators presented by Cohen and Levesque (1990):

$$\boldsymbol{Later}(\varphi) \stackrel{\text{def}}{=} \neg\varphi \wedge \boldsymbol{Poss}(\varphi)$$

$$\boldsymbol{Before}(\varphi, \psi) \stackrel{\text{def}}{=} (\forall x)(\boldsymbol{Af}(x, \psi) \rightarrow (\exists y)(\boldsymbol{Af}(y; x, \psi) \wedge \boldsymbol{Af}(x, \varphi)))$$

Event-based emotions depend on the *perception* of events. For this purpose, it is assumed that an agent perceived (*i.e.*, had the perception of) the event ε ($\boldsymbol{Percv}(\varepsilon)$) if it believes that this event happened immediately before now, which can be expressed in *AfPL* by the following definition:

$$\boldsymbol{Percv}(\varepsilon) \stackrel{\text{def}}{=} B(\boldsymbol{Last}(\varepsilon))$$

AfPL is not a multiagent logic, so all actions are performed by its sole agent. Actions also are events, so any action that was executed can be perceived by the agent exactly in the same way as any other event. Thus, to check if an action was performed by the agent, it is enough to verify if the action really happened and was perceived by the agent. This is done by the operator $\boldsymbol{Exec}(\varepsilon)$, defined as follows:

$$\boldsymbol{Exec}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{Act}(\varepsilon) \wedge \boldsymbol{Last}(\varepsilon) \wedge \boldsymbol{Percv}(\varepsilon)$$

AfPL has operators to represent two basic concepts of the BDI model: beliefs and goals. The *persistent goals* and related *intentions* operators, commonly used in BDI logics, are

defined by syntactic means. According to the analysis of Cohen and Levesque (1990) and van der Hoek and Wooldridge (2012), a *persistent goal* $\boldsymbol{PG}(\varphi)$ is a goal φ believed to be possible, but currently not satisfied, and a rational agent will hold the goal that φ be possible until it believes that it is true or that there is no way to accomplish it:

$$\boldsymbol{PG}(\varphi) \stackrel{\text{def}}{=} \boldsymbol{G}(\boldsymbol{Later}(\varphi)) \ \wedge \boldsymbol{B}(\neg\varphi) \ \wedge \ \boldsymbol{Before}(\boldsymbol{B}(\varphi) \vee \boldsymbol{B}(\neg\boldsymbol{Poss}(\varphi)), \neg\boldsymbol{G}(\boldsymbol{Later}(\varphi)))$$

A persistent goal $\boldsymbol{PG}(\varphi)$ indicates that the agent has a non-transitory interest to accomplish the goal, but it does not imply that the agent is actively working to do that. To start working on a persistent goal, an agent must have the intention to do so. According to Cohen and Levesque (1990), intentions can be divided into intentions to perform an action, represented in *AfPL* by the $\boldsymbol{IAct}(\varepsilon)$ operator, and intentions to reach a future state of the world, represented by the $\boldsymbol{ISt}(\varphi)$ operator.

The intention $\boldsymbol{IAct}(\varepsilon)$ of an agent to do a particular action ε implies that this agent is persistently aiming to execute the action, knowing when it is was executed (Cohen & Levesque, 1990; van der Hoek & Wooldridge, 2012):

$$\boldsymbol{IAct}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{Act}(\varepsilon) \wedge \boldsymbol{PG}(\boldsymbol{Next}(\varepsilon)) \wedge \boldsymbol{Bf}(\varepsilon, \boldsymbol{B}(\boldsymbol{Next}(\varepsilon)))$$

The knowledge that an action ε was executed is formalized by the belief, held immediately before the execution of ε, that this action would be executed immediately next.

An intention to reach a future state of the world is a more complex concept than an intention to perform an action. According to Cohen and Levesque (1990) and Wooldridge (2009, p. 68), this future directed kind of intention is assumed to: (a) drive means-end reasoning, posing planning problems believed to be solvable, (b) persist: they are persistent goals, (c) constrain future deliberation, influencing future actions and beliefs and (d) not be irrational: they must not conflict with other goals and intentions. In *AfPL* the $\boldsymbol{ISt}(\varphi)$ operator will represent intentions to bring about a future state φ of the world, allowing the definition of the *long term intentions* of the agent. The definition of $\boldsymbol{ISt}(\varphi)$ follows the $(INTEND_2\ x\ p)$ operator (Cohen & Levesque, 1990):

$$\boldsymbol{ISt}(\varphi) \stackrel{\text{def}}{=} \boldsymbol{PG}((\exists x)(\boldsymbol{Act}(x) \wedge \boldsymbol{Last}(x) \wedge \boldsymbol{Af}(x, \varphi) \wedge$$
$$\boldsymbol{Before}(\boldsymbol{B}((\exists y)\boldsymbol{Af}(y, \varphi)) \wedge \neg\boldsymbol{PG}(\neg\boldsymbol{Af}(x, \varphi), \boldsymbol{Last}(x)))$$

Note that, according to Cohen and Levesque (1990), to formalize intentions it is not necessary to define how the planning process works, but only to assume that as long as the agent believes there is a plan, the planning process is happening.

6. Modeling OCC Variables

Emotions in the OCC model depend on several cognitive *variables*, which are related to the mental states of the agent and its physiological state. These variables are instrumental in the process of experiencing emotions, because, as explained in 2.2, they determine whether an emotion will actually occur. Furthermore, these variables also determine the intensity of the emotions.

The OCC model divides these variables in central intensity variables, like *desirability*, and local variables, like *effort*. In our formal model, we need to define central variables such

as desirability and undesirability, as well as local variables such as likelihood and effort. All these variables, together with some accessory variables related to perception and action, are formally defined in the present section.

6.1 OCC Goals and BDI Intentions

Before we proceed to the formal definition of OCC model variables, we need to understand how goals in the OCC are related to goals in the BDI, defining clearly what an OCC goal means in the context of the BDI model. According to the OCC model, a goal is a kind of thing that someone desires to succeed: it is something the agent is *interested in* (Ortony et al., 1990, pp. 39-41), but the agent is not necessarily working actively to achieve it (these are the I goals in the OCC model terminology). Goals can also be a thing that the agent is *actively* pursuing, developing plans and acting to achieve it (Ortony et al., 1990, pp. 39-41) (A goals, in the OCC model terminology)[5].

The persistent goals $PG(\varphi)$ in *AfPL* are related to OCC's interest goals. These concepts are clearly interrelated: why the agent would persist with a goal when it is not interested in it?, or, conversely, why would we say that an agent is interested in a goal, if this "interest" lasts only for a very brief moment? Anyway, in our model we will equate these two concepts: $PG(\varphi)$ defines the OCC's interest goals of the agent.

To represent active-pursuit goals of the OCC model (A goals), we assume that it is necessary to consider the concept of intentions directed to future states of the world, which is used in the BDI model. The fact that an active-pursuit goal (an A goal in the OCC model) is something being actively pursued by the agent through some kind of plan clearly leads to the interpretation that A goals of the OCC model must match the notion of long term intentions $ISt(\varphi)$ of *AfPL*, which are the intentions to reach a future state of the world. The $PG(\varphi)$ persistent goals of *AfPL* cannot be considered equivalent to the OCC's A goals, because they do not support the notions of action and planning required by the A goals.

6.2 Desirability and Undesirability

Event-based emotions depend mainly on the desirability variable, although some global variables (*i.e.*, variables that affect the intensity of all emotions) can also have influence over them (see section 2.2). The desirability of an event is the degree to which the event appears to have beneficial consequences (*i.e.*, it is positively desirable) or as having harmful consequences (*i.e.*, it is negatively desirable or undesirable). As the desirability corresponds to a degree, it is important that it can be represented by logical language. For this reason, we believe that *AfPL* is an appropriate logic to formalize the OCC model.

According to the OCC model, in order to estimate the desirability or undesirability of an event, it is necessary to estimate (*i*) the importance of the goals affected by the event, and (*ii*) the possible degree to which the (consequences of the) event will advance or hinder the achievement of these goals.

---

5. The OCC model also defines a third type of goals: the replenishment (or repetitive) goals (R goals), which are goals that are not abandoned once they are achieved. In the present work, we are more interested in I and A goals. R goals can be represented in *AfPL*, however, for the sake of brevity, we do not explore this possibility in this paper.

From a formal point of view, the estimation of the importance of a goal φ depends on the desire force $\boldsymbol{df}(\varphi)$ assigned to this goal by the agent, be it a persistent goal $\boldsymbol{PG}(\varphi)$ or a long term intention $\boldsymbol{ISt}(\varphi)$. The desire force $\boldsymbol{df}(\varphi)$ will give all the information needed about the importance of the goal φ.

However, the estimation on how much an event ε advances (or hinders) the achievement of a goal φ is based, in the OCC model, on a hypothetical reasoning that depends on the consequences of the event ε, if it really happens. This work focuses on the effects of this event in the pursuing process of a particular goal, verifying whether the accomplishment of this goal is still possible. The modality operator for future possibility $\boldsymbol{Poss}(\varphi)$ will be used to verify wheter a goal φ is possible or not.

Now, the hypothetical condition of the reasoning can be represented by the subjective conditional probability on the possibility to accomplish the goal φ, given that the event ε will happen next: $\boldsymbol{bd}(\boldsymbol{Poss}(\varphi) \mid \boldsymbol{Next}(\varepsilon))$. The probability resulting from this conditional operator will give an estimation of how much probable (or possible) is the achievement of φ after the occurrence of ε.

By itself, this conditional probability does not indicate whether the occurrence of the event ε is advantageous, or unfavourable to the pursuit of φ. To have a measure that indicates whether the occurrence of the event ε is advancing (or hindering) the achievement of the goal φ, it is necessary to compare this hypothetical condition with the converse condition: $\boldsymbol{bd}(\boldsymbol{Poss}(\varphi) \mid \neg\boldsymbol{Next}(\varepsilon))$, which estimates what is the possibility to finish the goal φ if the event ε will not happen next.

To have an estimation of the amount of advancement or retrocession caused by the event ε, it suffices to look at the difference between these two conditional probabilities. This is estimated by the $\Delta_{des}(\varepsilon, \varphi)$ function defined as follows:

$$\Delta_{des}(\varepsilon, \varphi) \stackrel{\mathrm{def}}{=} \boldsymbol{bd}(\boldsymbol{Poss}(\varphi)|\boldsymbol{Next}(\varepsilon)) - \boldsymbol{bd}(\boldsymbol{Poss}(\varphi)|\neg\boldsymbol{Next}(\varepsilon))$$

A positive $\Delta_{des}(\varepsilon, \varphi)$ value indicates an advancement (a gain, or increase) and a negative value a retrocession (a loss, or decrease) in the chance to achieve φ.

The concept of *desirability* of the event ε with respect to the φ goal can be formalized as the combination of the importance (force) of the desire for φ and the amount of advancement brought by ε, if the event ε is really beneficial to the achievement of φ. In this work, we specify that the desirability of ε with respect to φ is given by the product of the desire force of φ and how much ε advances the realization of φ. This product indicates the positive gain resulting from the previous occurrence of the event, and this gain is equated in our formal OCC model to the desirability of the event. Thus, the $\boldsymbol{desblty}(\varepsilon, \varphi)$ function, which formalizes this concept, is defined for positive values of $\Delta_{des}(\varepsilon, \varphi)$ as follows:

$$\boldsymbol{desblty}(\varepsilon, \varphi) \stackrel{\mathrm{def}}{=} \boldsymbol{df}(\varphi) \times \Delta_{des}(\varepsilon, \varphi) \times \boldsymbol{?}(\Delta_{des}(\varepsilon, \varphi) > \boldsymbol{0})$$

Note that even using a product $\boldsymbol{df}(\varphi) \times \Delta_{des}(\varepsilon, \varphi)$, this definition of desirability is not related to the notion of motivation, which is loosely defined in the OCC model as the product of the desirability of an event and the subjective probability to achieve a goal (Ortony et al., 1990, p. 49). In our work, the desirability of an event with respect to a particular goal is the product of the desire force of the goal and the increase (gain) on the probability to achieve this goal caused by the event. The resulting product modulates the

desire force to achieve the goal by the increase on the chance to achieve it caused by the event. It is almost the opposite idea of motivation, and, to us, it captures the essence of the desirability concept.

The dual concept of undesirability is defined for negative values of $\Delta_{des}(\varepsilon, \varphi)$ in a similar way:

$$\boldsymbol{undesblty}(\varepsilon, \varphi) \stackrel{\text{def}}{=} \boldsymbol{df}(\varphi) \ \times \ -\Delta_{des}(\varepsilon, \varphi) \ \times \ \boldsymbol{?}(\Delta_{des}(\varepsilon, \varphi) < \boldsymbol{0})$$

The $\boldsymbol{undesblty}(\varepsilon, \varphi)$ function estimates the decrease in value (*i.e.*, the loss) caused by the event ε in the value of φ, calculated in terms of how much the chance to achieve φ decreases after ε happens.

6.3 Expectation, Prospect and Likelihood

In this work, the *expectation* that an event ε will occur in the near future ($\boldsymbol{Expec}(\varepsilon)$) is formalized as the belief that this event will happen next:

$$\boldsymbol{Expec}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{B}(\boldsymbol{Next}(\varepsilon))$$

In the OCC model, prospect-based emotions do not depend directly on the expectation about an event, but they depend on the *prospect* of the event. The notion of prospect is elementary in the OCC model, but it is not explicitly defined by Ortony et al. (1990). The common sense (*i.e.*, dictionary) definition of prospect says that a prospect is a mental picture or a vision about an event that is awaited or expected. According to this definition, there could be a subtle difference between prospect and expectation: a prospect would be a belief about another belief (*i.e.*, the expectation). This interpretation could lead to a formal definition of prospect as $\boldsymbol{B}(\boldsymbol{Expec}(\varepsilon))$. However, due to the KD45 modality of the $\boldsymbol{B}(\varphi)$ belief operator, this definition of prospect is equivalent to the definition of $\boldsymbol{Expec}(\varepsilon)$, thus in this work we will consider prospect and expectation as equivalent concepts and only use the $\boldsymbol{Expec}(\varepsilon)$ operator.

Now, let us analyse the concept of *likelihood* of an event. Following the discussion about likelihood in the OCC model (Ortony et al., 1990, p. 70), we assume that the likelihood of an event is not a necessarily conscious estimation about the probability of the occurrence of this event in the future. Thus, we define the likelihood for an event ε ($\boldsymbol{likelhd}(\varepsilon)$) as the probabilistic estimation that the event ε will happen in the future, formalized as follows:

$$\boldsymbol{likelhd}(\varepsilon) \stackrel{\text{def}}{=} \boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(\varepsilon)))$$

Note that being not conscious simply means that the agent does not necessarily believe that it estimated the likelihood of an event. To be conscious, the agent must believe it has this estimation, that is, it must have a belief similar to $\boldsymbol{B}(\boldsymbol{likelhd}(\varepsilon) = p)$.

6.4 Effort and Realization

Some prospect-based emotions depend on actions caused by agents. The intensity of these emotions depend on the *effort* expended by the agent in performing an action (Ortony et al., 1990, p. 71). The *AfPL* formal model for emotions considers only the *subjective* effort made to overcome contrary desires to not have performed an action. Thus, this effort is estimated

by the difference between the desire force to not have performed the action and the desire force to have performed the action. Note that the first desire is hypothetical, because the action was actually performed by the agent. So it is modeled by the belief of not having performed the action. If the desire force's difference is positive, then the desire not to have performed the action was greater than the desire to have done it, and some (subjective) effort was necessary to overcome the contrary desire. A positive difference provides an estimation of how much of this effort was made to have executed the action. This leads to the following definition of $\Delta_{eff}(\alpha)$, which estimates the effort to perform the action α based on the difference of the desire forces:

$$\Delta_{eff}(\alpha) \stackrel{\text{def}}{=} (df(B(\neg Exec(\alpha))) - df(Exec(\alpha))) \times ?(df(B(\neg Exec(\alpha))) \geq df(Exec(\alpha)))$$

Desire forces do not have a predefined maximum value, thus to normalize the values obtained with $\Delta_{eff}(\alpha)$ a normalization factor is necessary. This factor is $df(B(\neg Exec(\alpha))$ if it is effectively greater than $df(Exec(\alpha))$ (*i.e.*, the action α was actually an effort), otherwise it will be 1:

$$norm_{eff}(\alpha) \stackrel{\text{def}}{=} ?(df(B(\neg Exec(\alpha))) > df(Exec(\alpha))) \times df(B(\neg Exec(\alpha)))$$
$$+ ?(df(B(\neg Exec(\alpha))) \leq df(Exec(\alpha)))$$

Using $\Delta_{eff}(\alpha)$ and $norm_{eff}(\alpha)$, the effort to have performed the action α is estimated by the function $effort(\alpha)$, defined as follows:

$$effort(\alpha) \stackrel{\text{def}}{=} (\Delta_{eff}(\alpha) \div norm_{eff}(\alpha)) \times ?(Act(\alpha))$$

In *AfPL*, it is possible to model an epistemic version of the secondary variable *realization*, which represents the degree to which a confirmed or disconfirmed event was realized. Events in *AfPL* either occur or not, there is no "degree of realization" to be assigned to an event. However, it is possible to assign a probability to the *belief* that the event will happen next, implying that the "realization" of the event is completed. This probability offers an estimation of how much the realization of the event is "completed" on the basis that if someone is 100% certain that the event will happen next, then he or she believes that its realization is completed, otherwise if there is a less than 100% probability that the event will happen next, then this probability can be used as an estimation of how much still needs to be done. This leads to the following formalization for the estimation of the realization of an event:

$$realiz(\varepsilon) \stackrel{\text{def}}{=} bd(Next(\varepsilon)))$$

7. Modeling Event-Based Emotions with Consequences for Self

Event-based emotions are emotions that are elicited from the evaluation of the consequences of an event according to an agent's goals. If the event can be anticipated, it is a prospect-relevant emotion, otherwise it is prospect-irrelevant. This section presents the *AfPL* formalization of event-based emotions, which bring consequences to the agent (*i.e.*, the "self"), including prospect-relevant and prospect-irrelevant emotions, such as joy, distress, hope, fear, satisfaction, disappointment, relief and fears-confirmed.

7.1 Event-Based Emotions: Joy and Distress

According to the OCC model, *joy* and *distress* can occur when a person evaluates the desirability of an event in relation to the person's goals. Joy occurs when a person is pleased about a desirable event that has happened and distress occurs when a person is displeased about the happening of an undesirable event. The desirability of an event (*i.e.*, if it is desirable or undesirable) depends on how a person construes the consequences of an event as being desirable or undesirable. If an event promotes someone's goals, it is considered desirable by that person. Then, if the consequences of an event are desirable according to a person's goals (*i.e.*, it promotes the goals), then the person feels joy. On the other hand, he or she feels distress when the consequences of the event are evaluated as undesirable. For example, if a person learns that he or she is to receive a small and completely unexpected inheritance as the result of the death of a distant and unknown relative, this person would usually be pleased about receiving the inheritance.

The desirability is the main variable that determines the intensity of joy and distress emotions. In fact, there are global variables that determine the intensity of all emotions, such as sense of reality, proximity, and unexpectedness. However, we did not elaborate the representation of these other variables, because they require the representation of human physiology, which can not be expressed in AfPL. Furthermore, the OCC model authors state that the desirability is a central variable that determines the elicitation of joy and distress. Thus, to simplify the model, we use only desirability and undesirability variables, together with the perception operator defined in Section 6, in the formalization of the elicitation process of the intensity of the joy and distress emotions.

According to the OCC model, affective reactions are effectively experienced as emotions only when achieving a minimum intensity level. Affective reactions have only *potential* for the emotion, however, if this potential surpasses the minimum threshold, the emotion is experienced with an intensity related to the *potential* (Ortony et al., 1990, pp. 182-186). Because of this, we need some threshold level coefficients to formalize emotions. The coefficients min_{joy} and min_{distr} provide the necessary absolute threshold limits values for joy and distress, respectively. They represent the minimum intensity for an affective reaction to be experienced as an emotion of joy or distress. All other emotions have similar minimum constants.

The process described above is illustrated in Figure 2, which shows that the consequences of an actual event are evaluated according to a person's goals. The person is pleased when he or she evaluates the event or its consequences as desirable and displeased when they are evaluated as undesirable. A function of intensity assignment will define when the affective reaction of pleased or displeased will be experienced as an emotion and its intensity.

In the work that describes the OCC model (Ortony et al., 1990, pp. 182-186), there is a set of rules, similar to expert system rules, that are suggested as a computational tool to determine the potential and the intensity of several emotions, including joy, fear, and relief (see Figure 3 for an example of these rules). This set of rules does not define a formal model, because neither the semantics of the inference system nor the semantics of the elements (*i.e.*, predicates and functions) that appear on these rules are defined by Ortony et al. (1990). However, these rules can be used as a guide for the definition of *AfPl* numerical expressions, which will be able to calculate the potential and intensity of the OCC emotions. Due to
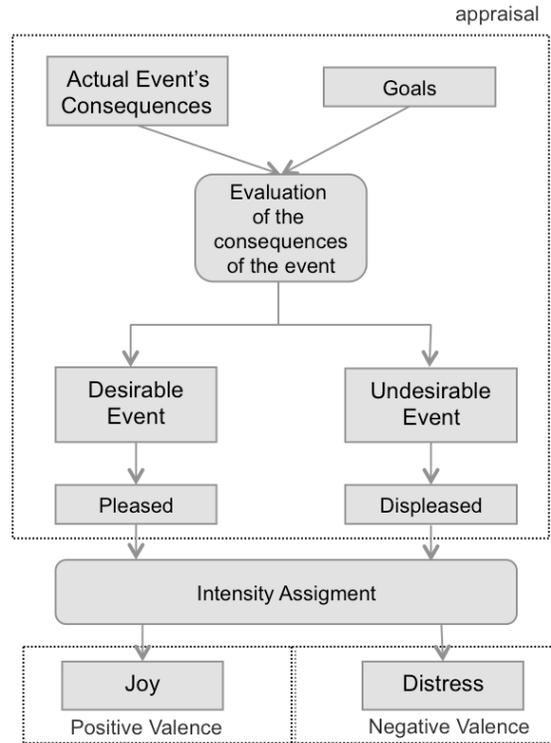
Figure 2: Scheme representing appraisal for joy and distress

the expressive power of *AfPl*, it is possible to accurately follow the initial attempts of a rule-based computational model for these emotions, as presented by Ortony et al. (1990, pp. 182-186), and define formal functions that represent the potential and the intensity of these emotions.

Event-based emotions depend on the consequences of a perceived event with respect to the goals of some agent (see Figure 2). The OCC model works with three kind of goals: interest, active-pursuit and replenishment goals. However, only two kind of goals are considered in the *AfPL* formalization of appraisal process: interest goals, which are represented by the BDI's persistent goals ($\boldsymbol{PG}(\varphi)$), and active-pursuit goals, which are mapped into the BDI's long term intentions ($\boldsymbol{ISt}(\varphi)$) (for more details, see Section 6.1). Thus, given an event ε perceived by the agent ($\boldsymbol{Percv}(\varepsilon)$) and a goal φ in which the agent is interested, or that the agent is actively pursuing, it is possible to start the appraisal process of emotions that depends on the consequences of ε over φ. If the degree of desirability of the event ε with respect to φ is greater than zero, then this event has desirable consequences for the goal φ. If the agent perceives ε, then the agent has the potential to feel joy with respect to φ. If the desirability is greater than zero (0), the potential for joy is equated to the desirability of the event ε with respect to φ. Otherwise, this potential is zero, meaning that the affective reaction will not be experienced as an emotion. This entails the following definition of the $\boldsymbol{joy\text{-}poten}(\varepsilon, \varphi)$ detection function:

$$\boldsymbol{joy\text{-}poten}(\varepsilon, \varphi) \stackrel{\text{def}}{=} \boldsymbol{desblty}(\varepsilon, \varphi) \times \boldsymbol{?}((\boldsymbol{PG}(\varphi) \vee \boldsymbol{ISt}(\varphi)) \wedge \boldsymbol{Percv}(\varepsilon) \wedge \boldsymbol{desblty}(\varepsilon, \varphi) > \boldsymbol{0})$$

Following the rules that evaluate the intensity of the emotion joy (Ortony et al., 1990, p. 183), we can specify that joy will only be felt if its potential surpasses the threshold for joy (*i.e.*, min_{joy}). If this is true, then the intensity of joy equals the difference between its potential and this threshold.

$$\boldsymbol{joy}(\varepsilon, \varphi) \stackrel{\text{def}}{=} (\boldsymbol{joy\text{-}poten}(\varepsilon, \varphi) - min_{joy}) \times \boldsymbol{?}(\boldsymbol{joy\text{-}poten}(\varepsilon, \varphi) > min_{joy})$$

Similarly to the case of joy, the functions $\boldsymbol{distr\text{-}poten}(\varepsilon, \varphi)$ and $\boldsymbol{distress}(\varepsilon, \varphi)$ represent, respectively, the potential and the intensity of distress with respect to the undesirable consequences of the event ε in relation to a goal φ. The potential for distress equals this negative desirability value if this event is perceived and its undesirability is greater than zero:

$$\boldsymbol{distr\text{-}poten}(\varepsilon, \varphi) \stackrel{\text{def}}{=} \boldsymbol{undesblty}(\varepsilon, \varphi) \times$$
$$\boldsymbol{?}((\boldsymbol{PG}(\varphi) \vee \boldsymbol{ISt}(\varphi)) \wedge \boldsymbol{Percv}(\varepsilon) \wedge \boldsymbol{undesblty}(\varepsilon, \varphi) > \boldsymbol{0})$$

As in the case of joy, the intensity of distress will only be felt if its potential surpasses the threshold level for distress min_{distr}:

$$\boldsymbol{distress}(\varepsilon, \varphi) \stackrel{\text{def}}{=} (\boldsymbol{distr\text{-}poten}(\varepsilon, \varphi) - min_{distr}) \times \boldsymbol{?}(\boldsymbol{distr\text{-}poten}(\varepsilon, \varphi) > min_{distr})$$

7.2 Prospect-Based Emotions: Hope and Fear

The pairs of emotions *hope/fear*, *satisfaction/disappointment*, and *fears-confirmed/relief* are prospect-relevant, event-based emotions. These emotions result from reacting to the prospect of desirable or undesirable events. They differ within the pairs based on the concrete occurrence or not of a particular event. For instance, when a student accomplishes an exam, he or she has the *hope* to obtain a high score, or equivalently he or she has the prospect to obtain a high score. To obtain a high score is an event that has not happened yet, but for which the student hopes. The student will be *satisfied* if his or her scores are, indeed, high, or *disappointed* otherwise (see Figure 4).

As in the case of prospects irrelevant emotions, the formalization of these emotions will consider the set of rules for fear and relief presented by Ortony et al. (1990, pp. 185-186) and define detecting functions for the potential and the intensity of each one of these emotions.

The formal analysis starts with the emotions hope and fear. When someone expects that an event will occur, but it has not happened yet, then the person experiences hope or fear depending upon, respectively, whether the event is desirable or not. For the example case of the student who recently took a test, the student experiences hope until he or she knows the result of the exam.

The potential and the intensity of the fear emotion are detected by the functions $\boldsymbol{fear\text{-}poten}(\varepsilon, \varphi)$ and $\boldsymbol{fear}(\varepsilon, \varphi)$. The potential function is defined over the prospective expectation and desirability of the event ε in respect to a goal φ, as follows:

$$\boldsymbol{fear\text{-}poten}(\varepsilon,\varphi) \stackrel{\text{def}}{=} \boldsymbol{undesblty}(\varepsilon,\varphi) \times \boldsymbol{likelhd}(\varepsilon) \times$$
$$\boldsymbol{?}((\boldsymbol{PG}(\varphi) \vee \boldsymbol{ISt}(\varphi)) \wedge \boldsymbol{Expec}(\varepsilon) \wedge \boldsymbol{undesblty}(\varepsilon,\varphi) > \boldsymbol{0})$$

In the case of prospective relevant emotions, the likelihood of the occurrence of the event becomes an important factor on the determination of the potential of the emotion. The rules defined by Ortony et al. (1990, p. 185) assume that an unspecified function combines undesirability and likelihood. But, in the *AfPL* formal model, the likelihood variable works as a linear amplifier of the undesirability. The value of the undesirability is proportional to the chance of the occurrence of the event.

The fear emotion will only be felt if its potential surpasses the threshold level for fear. Thus, the intensity of fear is determined by the following function:

$$\boldsymbol{fear}(\varepsilon,\varphi) \stackrel{\text{def}}{=} (\boldsymbol{fear\text{-}poten}(\varepsilon,\varphi) - min_{fear}) \times \boldsymbol{?}(\boldsymbol{fear\text{-}poten}(\varepsilon,\varphi) > min_{fear})$$

The formal model of the emotion hope is very similar to the model of fear. The potential and intensity of the emotion hope are calculated, respectively, by the $\boldsymbol{hope\text{-}poten}(\varepsilon,\varphi)$ and $\boldsymbol{hope}(\varepsilon,\varphi)$ functions, defined as follows:

$$\boldsymbol{hope\text{-}poten}(\varepsilon,\varphi) \stackrel{\text{def}}{=} \boldsymbol{desblty}(\varepsilon,\varphi) \times \boldsymbol{likelhd}(\varepsilon) \times$$
$$\boldsymbol{?}((\boldsymbol{PG}(\varphi) \vee \boldsymbol{ISt}(\varphi)) \wedge \boldsymbol{Expec}(\varepsilon) \wedge \boldsymbol{desblty}(\varepsilon,\varphi) > \boldsymbol{0})$$
$$\boldsymbol{hope}(\varepsilon,\varphi) \quad \stackrel{\text{def}}{=} \quad (\boldsymbol{hope\text{-}poten}(\varepsilon,\varphi) - min_{hope}) \times \boldsymbol{?}(\boldsymbol{hope\text{-}poten}(\varepsilon) > min_{hope})$$

7.3 Prospect-Based Emotions: Relief and Fears-Confirmed

If a prospective undesirable event is disconfirmed, then someone can experience an emotion of relief. For example, if a student is expected to have a low score in an exam, he or she could experience relief when realizing that the grade was not so bad as expected.

Similarly to the case of joy and fear, the set of rules defined for the relief emotion (Ortony et al., 1990, p. 186) helped with the definition of this function. Figure 3 shows the original rules defined in the OCC model for the elicitation of the potential and intensity of the relief emotion. The potential for relief is estimated based on the previous fear potential for the event (FEAR-POTENTIAL(p, e, t)), the degree on the realization of this event (REALIZATION(e, t_2)), the effort to have avoided it (EFFORT(p, e)) and an unspecified global intensity factor ($I_g(p, e, t_2)$), which are combined by an unspecified $f_r[...]$ function.

The formalization of the emotions relief and fears-confirmed is focused on the variables effort and realization, assuming that the relief is proportional to these variables. The global intensity factor ($I_g(p, e, t_2)$) is not explicitly defined in the *AfPL* formalization. However, it can be applied to estimation functions of intensity of emotions, as a global multiplier factor used to amplify or attenuate the intensity of an emotion.

Effort can be a significant factor for relief and fears-confirmed emotions, but only if some effort was effectively spent on avoiding the event, implying in a plan or sequence of actions α being performed by the agent with this purpose. *AfPL* does not have a *null* or *no operation* action, so two sets of potential and intensity functions were defined to handle this situation.

(5) IF FEAR-POTENTIAL$(p, e, t) > 0$ AND DISBELIEVE(p, e, t_2) AND $t_2 \geqslant t$
    THEN set RELIEF-POTENTIAL$(p, e, t_2) =$
        f_r[FEAR-POTENTIAL(p, e, t), EFFORT(p, e),
        REALIZATION(e, t_2), $I_g(p, e, t_2)$]
(6) IF RELIEF-POTENTIAL$(p, e, t_2) >$ RELIEF-THRESHOLD(p, t_2)
    THEN set RELIEF-INTENSITY$(p, e, t_2) =$
        RELIEF-POTENTIAL(p, e, t_2) - RELIEF-THRESHOLD(p, t_2)
    AND reset FEAR-POTENTIAL$(p, e, t_2) =$
        f_f[| DESIRE(p, e, t_2) |, LIKELIHOOD(p, e, t_2), $I_g(p, e, t_2)$]
    ELSE set RELIEF-INTENSITY$(p, e, t_2) = 0$

Figure 3: OCC model rules to calculate relief (Ortony et al., 1990, p. 186)

The **relief-poten**(ε, φ) function models the determination of relief potential when no effort (and thus no sequence of actions α) was made to avoid the event ε. It is defined as follows:

$$\textbf{relief-poten}(\varepsilon, \varphi) \stackrel{\text{def}}{=} \textbf{fear-poten}(\varepsilon, \varphi) \ \times \ (\textbf{1} + \textbf{realiz}(\varepsilon)) \ \times$$
$$\textbf{?}((\textbf{PG}(\varphi) \vee \textbf{ISt}(\varphi)) \wedge \neg\textbf{Percv}(\varepsilon) \wedge \textbf{Poss}(\textbf{B}(\neg\textbf{Next}(\varepsilon))) \wedge \textbf{fear-poten}(\varepsilon, \varphi) > \textbf{0})$$

The agent has a potential to feel relief if it disbelieves that this event happened and will occur $(\neg\textbf{Percv}(\varepsilon) \wedge \textbf{Poss}(\textbf{B}(\neg\textbf{Next}(\varepsilon))))$, even if there still is a potential to be fearful in respect to the event $(\textbf{fear-poten}(\varepsilon, \varphi) > \textbf{0}))$. The potential for relief is amplified by how much the agent believed that the event ε was close to happen $(\textbf{1} + \textbf{realiz}(\varepsilon))$.

The condition $\textbf{Poss}(\textbf{B}(\neg\textbf{Next}(\varepsilon)))$ is a formalization of the DISBELIEVE(p, e, t_2) predicate, which is true when, at future time t_2, the agent p believes that the event e will not occur (Ortony et al., 1990, p. 186). Note that at the present instant of time, the agent can still believe in the prospect of ε, implying that non-zero values for the **fear-poten**(ε, φ) and **realiz**(ε) functions are still possible.

The **relief**(ε, φ) function detects the intensity of the relief of the emotion when there is no effort involved and the potential for relief achieves a threshold.

$$\textbf{relief}(\varepsilon, \varphi) \stackrel{\text{def}}{=} (\textbf{relief-poten}(\varepsilon, \varphi) - min_{relief}) \times \textbf{?}(\textbf{relief-poten}(\varepsilon, \varphi) > min_{relief})$$

The "reset" command of fear potential executed in the OCC rules is not necessary in the *AfPL* formal model. Assuming $\textbf{Poss}(\textbf{B}(\neg\textbf{Next}(\varepsilon)))$, we have, by the definition of $\textbf{Poss}(\varphi)$, $(\exists x)\textbf{Af}(x, \textbf{B}(\neg\textbf{Next}(\varepsilon)))$, implying that there is a future instant of time x when $\textbf{B}(\neg\textbf{Next}(\varepsilon))$ is true. At this moment, by properties D and 5 from the KD45 modality, it will be possible to infer $\neg\textbf{B}(\textbf{B}(\textbf{Next}(\varepsilon)))$, which is equivalent to $\neg\textbf{Expec}(\varepsilon)$, setting the fear potential to zero.

The **relief-poten**$(\varepsilon, \varphi, \alpha)$ and **relief**$(\varepsilon, \varphi, \alpha)$ functions estimate, respectively, the relief potential and intensity when some effort was made to prevent the event ε to happen. In this case, the sequence of actions α represents the actions performed for this purpose.

$$\textbf{relief-poten}(\varepsilon, \varphi, \alpha) \stackrel{\text{def}}{=} \textbf{fear-poten}(\varepsilon, \varphi) \ \times \ (\textbf{1} + \textbf{effort}(\alpha)) \ \times (\textbf{1} + \textbf{realiz}(\varepsilon)) \ \times$$
$$\textbf{?}((\textbf{PG}(\varphi) \vee \textbf{ISt}(\varphi)) \wedge \textbf{Act}(\alpha) \wedge \neg\textbf{Percv}(\varepsilon) \wedge \textbf{Poss}(\textbf{B}(\neg\textbf{Next}(\varepsilon))) \wedge \textbf{fear-poten}(\varepsilon, \varphi) > \textbf{0})$$

$$\textbf{\textit{relief}}(\varepsilon, \varphi, \alpha) \overset{\text{def}}{=} (\textbf{\textit{relief-poten}}(\varepsilon, \varphi, \alpha) - min_{relief}) \times \textbf{?}(\textbf{\textit{relief-poten}}(\varepsilon, \varphi, \alpha) > min_{relief})$$

The fears-confirmed emotion is triggered by the perception of an expected undesirable event, confirming the fear felt about this event. The belief that the event is close to happen is a factor that aggravates (*i.e.*, amplifies) the fears-confirmed emotion. However, because the event was effectively perceived, it is possible to assume the maximum value to the secondary variable realization, which is a probabilistic variable. The effort devoted to prevent this event is also a factor that amplifies the potential of this emotion. Functions to estimate the potential and intensity of fears-confirmed taking into account the effort to avoid the fearful event (which is represented by the sequence of actions α), are defined as follows.

$$\textbf{\textit{fearconf-poten}}(\varepsilon, \varphi, \alpha) \overset{\text{def}}{=} \textbf{\textit{fear-poten}}(\varepsilon, \varphi) \; \times \; (\textbf{1} + \textbf{\textit{effort}}(\alpha)) \; \times (\textbf{1} + \textbf{\textit{realiz}}(\varepsilon)) \; \times$$
$$\textbf{?}((\textbf{\textit{PG}}(\varphi) \vee \textbf{\textit{ISt}}(\varphi)) \wedge \textbf{\textit{Act}}(\alpha) \wedge \textbf{\textit{Percv}}(\varepsilon) \wedge \textbf{\textit{fear-poten}}(\varepsilon, \varphi) > \textbf{0})$$
$$\textbf{\textit{fears-confirmed}}(\varepsilon, \varphi, \alpha) \overset{\text{def}}{=} (\textbf{\textit{fearconf-poten}}(\varepsilon, \varphi, \alpha) - min_{fearconf}) \times$$
$$\textbf{?}(\textbf{\textit{fearconf-poten}}(\varepsilon, \varphi, \alpha) > min_{fearconf})$$

The $\textbf{\textit{fears-confirmed}}(\varepsilon, \varphi)$ and $\textbf{\textit{fearconf-poten}}(\varepsilon, \varphi)$ functions that handle the case where no effort was put in preventing the event can be easily derived from the above definitions by eliminating terms that refer to α sequence of actions.

7.4 Prospect-Based Emotions: Satisfaction and Disappointment

The emotions of satisfaction and disappointment also depend on the confirmation or disconfirmation of the prospect of an event, but this time a desirable event (Ortony et al., 1990, p. 20). In our example, this means that the student knows whether a prospective desirable event will happen or not . The emotion satisfaction is elicited when one is pleased about the confirmation of the prospect of a desirable event, and disappointment is elicited when one is displeased about the disconfirmation of the prospect of a desirable event. Figure 4 presents a scheme that illustrates the appraisal for Satisfaction and Disappointment emotions.

For example, a student that receives his or her assessment scores experiences satisfaction if the scores are high (*i.e.*, by confirmation of the prospect of a desirable event) or disappointment if the student realizes that he or she was wrong about his or her first impression of the exam and in fact the score received was low.

The formal model to detect the emotion satisfaction is incorporated into the functions $\textbf{\textit{satisf-poten}}(\varepsilon, \varphi, \alpha)$ and $\textbf{\textit{satisf-inten}}(\varepsilon, \varphi, \alpha)$, defined as follows. The parameter α is a sequence of actions that represents the effort to make the event ε happen. Similarly to fears-confirmed, because the event was perceived, it is possible to assume that the probabilistic realization variable reached the maximum value.

$$\textbf{\textit{satisf-poten}}(\varepsilon, \varphi, \alpha) \overset{\text{def}}{=} \textbf{\textit{hope-poten}}(\varepsilon, \varphi) \; \times \; (\textbf{1} + \textbf{\textit{effort}}(\alpha)) \; \times (\textbf{1} + \textbf{\textit{realiz}}(\varepsilon)) \; \times$$
$$\textbf{?}((\textbf{\textit{PG}}(\varphi) \vee \textbf{\textit{ISt}}(\varphi)) \wedge \textbf{\textit{Act}}(\alpha) \wedge \textbf{\textit{hope-poten}}(\varepsilon, \varphi) > \textbf{0})$$
$$\textbf{\textit{satisfaction}}(\varepsilon, \varphi, \alpha) \overset{\text{def}}{=} (\textbf{\textit{satisf-poten}}(\varepsilon, \varphi, \alpha) - min_{satisf}) \times$$
$$\textbf{?}(\textbf{\textit{satisf-poten}}(\varepsilon, \varphi, \alpha) > min_{satisf})$$

The disappointment emotion is formalized by the potential and intensity functions defined below.
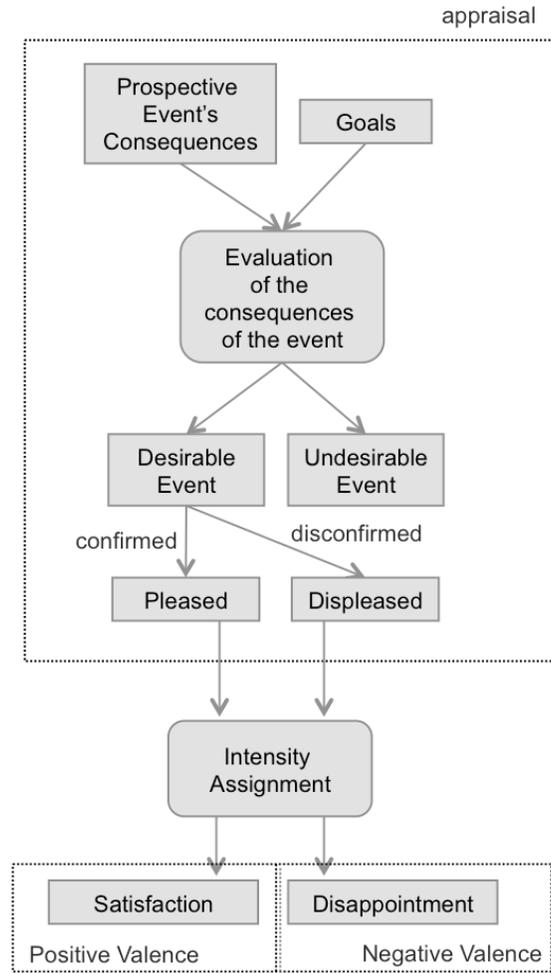
Figure 4: Scheme representing appraisal for satisfaction and disappointment

$$disapp\text{-}poten(\varepsilon, \varphi, \alpha) \stackrel{\text{def}}{=} hope\text{-}poten(\varepsilon, \varphi) \ \times \ (1 + effort(\alpha)) \ \times (1 + realiz(\varepsilon)) \ \times$$
$$?((PG(\varphi) \vee ISt(\varphi)) \wedge Act(\alpha) \wedge \neg Percv(\varepsilon) \wedge Poss(B(\neg Next(\varepsilon))) \wedge hope\text{-}poten(\varepsilon, \varphi) > 0)$$
$$disappointment(\varepsilon, \varphi, \alpha) \stackrel{\text{def}}{=} (disapp\text{-}poten(\varepsilon, \varphi, \alpha) - min_{disapp}) \times$$
$$?(disapp\text{-}poten(\varepsilon, \varphi, \alpha) > min_{disapp})$$

8. A Scenario

In this section, we present a scenario to illustrate the process of appraisal that happens to an agent, which has an emotional BDI architecture according to our proposed formal model of emotions. In this illustration, we consider the previously explained example of

the vacuum cleaner robot and its environment of a grid with two cells (see Section 3)[6]. Besides the previously cited desires (*i.e.*, recharge, clean the cell, or move to the next cell), our agent, which is called Nicky, also has the desire of protecting its own existence with a stronger priority.

While Nicky is cleaning Cell B, Pat, a clumsy researcher who also works in the same laboratory, goes in Nicky's direction without noticing it. When Nicky perceives that Pat is going to stomp on it, the desire of self-protection becomes an intention and Nicky feels fear of being damaged.

In order to alert the awkward scientist, Nicky emits an audible alarm. Pat perceives Nicky (almost behind her feet) and step on the other cell, avoiding trampling over Nicky. When Nicky perceives that it is not in danger any more, it finally feels relief. It can continue to do its work; at least until Pat decides to come back to her office.

Let us now see the process of appraisal in the *AfPL* formalization, when considered under the perspective of the reasoning cycles of the BDI architecture.

8.1 Initial Mental State

Nicky has the goal of cleaning the cells as its main intention. We can also expect Nicky to have basic self-preservation instinct, *i.e.*, it has the intention of preserving its own existence. Thus, Nicky's motivational state registers two main intentions:

$$\boldsymbol{ISt}(cleaning), \quad \boldsymbol{ISt}(selfpres) \tag{3}$$

Nicky's behaviour is driven by its emotions, so Nicky must have an estimation of how good it is to work, and to continue to exist, combined with thresholds for positive (joy, hope, etc.) and negative (distress, fear, etc.) emotions[7]:

$$\boldsymbol{df}(cleaning) = \boldsymbol{80}, \quad \boldsymbol{df}(selfpres) = \boldsymbol{100}$$
$$min_{joy} = \boldsymbol{50}, \quad min_{fear} = \boldsymbol{50} \tag{4}$$

Nicky's planning knowledge to achieve the intentions defined in (3) is driven by possible emotional responses caused by perceived or prospected events from the environment. This knowledge is specified by the following set of non-logical axioms:

$$\boldsymbol{ISt}(selfpres) \ \wedge \ \boldsymbol{ISt}(cleaning) \ \wedge \boldsymbol{joy}(dirty, cleaning) > \boldsymbol{0}$$
$$\wedge \ \boldsymbol{fear}(smash, selfpres) = \boldsymbol{0} \ \rightarrow \ \boldsymbol{IAct}(cleaning)$$
$$\boldsymbol{ISt}(selfpres) \ \wedge \ \boldsymbol{ISt}(cleaning) \ \wedge \ \boldsymbol{joy}(dirty, cleaning) = \boldsymbol{0} \tag{5}$$
$$\wedge \ \boldsymbol{fear}(smash, selfpres) = \boldsymbol{0} \ \rightarrow \ \boldsymbol{IAct}(move)$$
$$\boldsymbol{ISt}(selfpres) \ \wedge \boldsymbol{fear}(smash, selfpres) > \boldsymbol{0} \ \rightarrow \ \boldsymbol{IAct}(alarm)$$

Emotions provide motivations or restraints to the planning process which transforms Nicky's intentions into acts. Note, however, that this is an example designed for illustration

---

6. This scenario was previously presented by Gluz and Jaques (2014) for an example of implementation of BDI agents with emotions.

7. In the text following, belief degrees and desire forces are defined using a compact form, which follows the usual decimal notation for rational values. This is a convenient notation that can be translated to the standard syntax of *AfPL*. For instance, $\boldsymbol{bd}(cleaning) = \boldsymbol{0.8}$ is equivalent to $\boldsymbol{bd}(cleaning) = \boldsymbol{4/5}$, and then to $\boldsymbol{bd}(cleaning) \times \boldsymbol{5} = \boldsymbol{4}$. Integer constants like 4 or 5 are only the compact form of the sums of 1 constants, thus the standard form for the example is $\boldsymbol{bd}(cleaning) \times (\boldsymbol{1}+\boldsymbol{1}+\boldsymbol{1}+\boldsymbol{1}+\boldsymbol{1}) = \boldsymbol{1}+\boldsymbol{1}+\boldsymbol{1}+\boldsymbol{1}$.

purposes. It considers only one motivational emotion, "joy", to find dirt and start to work, and one restraint emotion, the "fear" of the agent to be smashed. In order to avoid to be smashed, and according to the fact that alarms are useful to prevent being smashed, Nicky plans to emit an alarm, expecting to receive new perception from its sensors informing that the possible cause of problem disappeared. If desired, it is possible to generalize this planning knowledge to a rule based system, letting events and active-pursuit goals as parameters of these plans.

Nicky also has some beliefs about what usually happens with dirty cells and what situations can damage the agent or not. Nicky believes that human approximation can damage a robot, because the human can trample on the robot, and also believes that if the human is departing then it should not expect to be damaged:

$$
\begin{aligned}
&\boldsymbol{B}(\boldsymbol{Last}(dirty) \rightarrow \boldsymbol{Next}(dirty)) \\
&\boldsymbol{B}(\boldsymbol{Last}(huApproach) \rightarrow \boldsymbol{Expect}(smash)) \\
&\boldsymbol{B}(\boldsymbol{Last}(huDepart) \rightarrow \neg\boldsymbol{Expect}(smash))
\end{aligned} \tag{6}
$$

To discover if Nicky will feel joy and start cleaning, we need to express the knowledge of how much the fact of finding dirt will advance the goal of cleaning. The following probabilistic model allows Nicky to estimate the desirability to do the cleaning, given it found dirt. It assigns conditional probabilities to the possibility to do some cleaning if dirt is found (1.0 if this is true, 0.3 otherwise). Note that this and the next probabilistic models are simple models specified for illustration purposes. However, if desired, it is possible to generalize these probabilistic models, from a simple conditional probability to a realistic bayesian network.

$$
\begin{aligned}
&\boldsymbol{bd}(\boldsymbol{Poss}(cleaning) \mid \boldsymbol{Next}(dirty)) = \mathbf{1} \\
&\boldsymbol{bd}(\boldsymbol{Poss}(cleaning) \mid \neg\boldsymbol{Next}(dirty)) = \mathbf{0.3}
\end{aligned} \tag{7}
$$

Nicky's general knowledge about how much to be smashed will hinder its self preservation is defined by the following probabilistic model:

$$
\begin{aligned}
&\boldsymbol{bd}(\boldsymbol{Poss}(selfpres) \mid \boldsymbol{Next}(smash)) = \mathbf{0.1} \\
&\boldsymbol{bd}(\boldsymbol{Poss}(selfpres) \mid \neg\boldsymbol{Next}(smash)) = \mathbf{1.0}
\end{aligned} \tag{8}
$$

The smash event is a very harmful event, thus Nicky must be able to make an assessment on the likelihood of this event, given observable evidences on the prospect of this event ($\boldsymbol{B}(\boldsymbol{Last}(huApproach))$). The probabilistic model presented below shows how this assessment is made:

$$
\begin{aligned}
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash))|\boldsymbol{Last}(huApproach)) = \mathbf{0.8} \\
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash))|\neg\boldsymbol{Last}(huApproach)) = \mathbf{0.1}
\end{aligned} \tag{9}
$$

8.2 First BDI Reasoning Cycle

Initially, Nicky is located in the cell B and perceives that this cell is dirty:

$$
\boldsymbol{Percv}(located\_cell\_b), \quad \boldsymbol{Percv}(dirty) \tag{10}
$$

Given these perceptions, it is not possible to infer $\boldsymbol{Expec}(smash)$, which implies that $\boldsymbol{fear\text{-}poten}(smash, selfpres) = \boldsymbol{0}$ (see Subsection 7.2) and thus:

$$\boldsymbol{fear}(smash, selfpres) = \boldsymbol{0}$$

As a consequence, Nicky feels no fear to be smashed. But, from the probabilistic model defined in (7), it is possible to estimate the value of $\Delta_{des}(dirty, cleaning)$:

$$
\begin{aligned}
\Delta_{des}&(dirty, cleaning) \\
&= \boldsymbol{bd}(\boldsymbol{Poss}(cleaning)|\boldsymbol{Next}(dirty)) - \boldsymbol{bd}(\boldsymbol{Poss}(cleaning)|\neg\boldsymbol{Next}(dirty)) \\
&= \boldsymbol{1 - 0.3 = 0.7}
\end{aligned}
$$

Now, using the value defined in (4) for $\boldsymbol{df}(cleaning)$, it is possible to calculate the desirability of the dirty event for the cleaning goal:

$$
\begin{aligned}
\boldsymbol{desblty}&(dirty, cleaning) \\
&= \boldsymbol{df}(cleaning) \ \times \ \Delta_{des}(dirty, cleaning) \ \times \ \boldsymbol{?}(\Delta_{des}(dirty, cleaning) > \boldsymbol{0}) \qquad (11) \\
&= \boldsymbol{80 \times 0.7 \times} \ \boldsymbol{?(0.7 > 0) =} \ \boldsymbol{80 \times 0.7 \times} \ \boldsymbol{1} \ = \boldsymbol{56}
\end{aligned}
$$

With (11), (4), and (10), it is possible to infer $\boldsymbol{joy\text{-}poten}(dirty, cleaning) = \boldsymbol{56}$, which implies $\boldsymbol{joy}(dirty, cleaning) = \boldsymbol{6}$ (see Subsection 6.2), and thus $\boldsymbol{joy}(dirty, cleaning) > \boldsymbol{0}$, so Nicky will be happy to start cleaning the floor:

$$\boldsymbol{IAct}(clean)$$

8.3 Second BDI Reasoning Cycle

In the second cycle, while cleaning the perceived dirt, Nicky receives from its vision sensor the perception "Human approaching":

$$\boldsymbol{Percv}(dirty), \quad \boldsymbol{Percv}(huApproach)$$

This, by the definition of $\boldsymbol{Percv}(\varepsilon)$ (see Subsection 5.3), is equivalent to:

$$\boldsymbol{B}(\boldsymbol{Last}(huApproach)) \qquad (12)$$

Now, with (6), the transitivity of the $\mathcal{B}$ relation (see Subsection 5.2), and some propositional inferences, Nicky's beliefs entail the expectation that Nicky will be smashed $\boldsymbol{B}(\boldsymbol{Expect}(smash))$, which is equivalent to say Nicky has the expectation to be smashed (see Section 6.3):

$$\boldsymbol{Expec}(smash)$$

This is one of the conditions to feel fear to be smashed. The other conditions (see Subsection 7.2) are related to the undesirability and likelihood of this event. The conditional probabilities defined in (8) allow Nicky to estimate the undesirability of being smashed:

$$
\begin{aligned}
\boldsymbol{undesblty}&(smash, selfpres) \\
&= \boldsymbol{df}(selfpres) \ \times \ -\Delta_{des}(smash, selfpres) \ \times \ \boldsymbol{?}(\Delta_{des}(smash, selfpres) < \boldsymbol{0}) \\
&= \boldsymbol{100 \times -(0 - 0.9) \times} \ \boldsymbol{?((0 - 0.9) < 0)} \\
&= \boldsymbol{100 \times 0.9 \times 1 =} \ \ \boldsymbol{90}
\end{aligned}
\qquad (13)
$$

The probabilistic model defined in (9) allows Nicky to estimate the likelihood to be smashed, given the prospect of this event. The probability spaces $\mathcal{P}_w = \langle \mathcal{W}_w, \mathcal{F}_w, \mu_w \rangle$ employed in *AfPL* models imply in the validity of the following equation[8]:

$$
\begin{aligned}
\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash))) &= \\
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \boldsymbol{Last}(huApproach)) + \\
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \neg\boldsymbol{Last}(huApproach))
\end{aligned}
\tag{14}
$$

Using basic bayesian reasoning, which is also valid in *AfPL*'s probability spaces, it is possible to derive the following equations:

$$
\begin{aligned}
\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \boldsymbol{Last}(huApproach)) &= \\
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash))|\boldsymbol{Last}(huApproach)) \times \boldsymbol{bd}(\boldsymbol{Last}(huApproach)) \\
\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \neg\boldsymbol{Last}(huApproach)) &= \\
&\boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash))|\neg\boldsymbol{Last}(huApproach)) \times \boldsymbol{bd}(\neg\boldsymbol{Last}(huApproach))
\end{aligned}
\tag{15}
$$

Given (12), it is also valid to assume, in *AfPL*'s probability spaces, that:

$$
\boldsymbol{bd}(\boldsymbol{Last}(huApproach)) = \boldsymbol{1}
\tag{16}
$$

Now, with (9), (14), (15), and (16), Nicky can calculate the likelihood to be smashed:

$$
\begin{aligned}
\boldsymbol{likelihood}(smash) &\\
&= \boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \\
&= \boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \boldsymbol{Last}(huApproach)) + \\
&\quad \boldsymbol{bd}(\boldsymbol{Poss}(\boldsymbol{Next}(smash)) \wedge \neg\boldsymbol{Last}(huApproach)) \\
&= \boldsymbol{0.8} \times \boldsymbol{1} + \boldsymbol{0.1} \times \boldsymbol{0} = \boldsymbol{0.8}
\end{aligned}
\tag{17}
$$

The potential and intensity of fear for Nicky's well being if it is smashed can be estimated based on the undesirability (13) and likelihood (17) of this event:

$$
\begin{aligned}
\boldsymbol{fear\text{-}poten}(smash, selfpres) &\\
&= \boldsymbol{undesblty}(smash, selfpres) \times \boldsymbol{likelhd}(smash) \times \\
&\quad \boldsymbol{?}(\boldsymbol{ISt}(\varphi) \wedge \boldsymbol{Expec}(smash) \wedge \boldsymbol{undesblty}(smash, selfpres) > \boldsymbol{0}) \\
&= \boldsymbol{90} \times \boldsymbol{0.8} \times \boldsymbol{1} = \boldsymbol{72} \\
\boldsymbol{fear}(smash, selfpres) &\\
&= (\boldsymbol{fear\text{-}poten}(smash, selfpres) - min_{fear}) \times \\
&\quad \boldsymbol{?}(\boldsymbol{fear\text{-}poten}(smash, selfpres) > min_{fear}) \\
&= (\boldsymbol{72} - \boldsymbol{50}) \times \boldsymbol{1} = \boldsymbol{22}
\end{aligned}
$$

Nicky's planning knowledge, as stated in (5), works only with three possible courses of action: to clean the floor, to sound the alarm, or to move and find new dirt. However, the fact that Nicky feels fear ($\boldsymbol{fear}(smash, selfpres) > \boldsymbol{0}$) and expects to be smashed show that the only course of action is to sound the alarm:

$$
\boldsymbol{IAct}(alarm)
$$

Note that, because $\boldsymbol{fear}(smash, selfpres) > \boldsymbol{0}$, Nicky will not perform any cleaning or moving action for now (it is not possible to satisfy the $\boldsymbol{fear}(smash, selfpres) = \boldsymbol{0}$ on cleaning plan).

---

8. A generalized version of this equation is, actually, a probabilistic axiom of *AfPL*.

8.4 Third BDI Reasoning Cycle

In the third cycle, let us assume that the alarm was effective and Pat, the researcher, heard it, and moved out of Nicky's way. At the same time, Nicky also perceived that the floor on cell B is not dirty any more. Thus, Nicky's perception base only contains:

$$\boldsymbol{Percv}(huDepart)$$

Based on this perception and (6), Nicky cannot hold the prospect to be smashed, and then it will not experience fear any more, because $\boldsymbol{fear\text{-}inten}(smash, selfpres)$ is reduced to zero. However, because $\boldsymbol{joy\text{-}inten}(dirty, cleaning)$ is also zero, Nicky cannot return to its cleaning activities. Given the planning knowledge specified in (5), the only possible course of action for Nick is to start to move and find a dirty cell:

$$\boldsymbol{IAct}(move)$$

9. Conclusions

In this paper, we presented a formalization of the appraisal process of emotions in a probabilistic BDI logic, known as *AfPL* Logic. Specifically, we modeled event-based emotions with consequences for self based on the OCC cognitive psychological theory of emotions. As already pointed out by Adam et al. (2009), a formal logic provides a universal vocabulary with clear semantics, and it enables the reasoning about and explanation of an agent's behavior. The proposed logical formalization has the advantage of being unambiguous and to allow analysis and verification. Besides, it formalizes the process of appraisal that depends on agents' beliefs, goals and desirability, a new concept introduced in the model proposed in this paper.

One important aspect of our formalization is that it allows us to compute the potential of an affective reaction triggered in the appraisal process by using a probabilistic logic. The amount of potential determines whether an affective reaction is actually experienced as an emotion or not. An affective reaction is only experienced as an emotion by a person if its potential achieves an emotion-specific threshold. Furthermore, this potential also allows the determination of the intensity of emotions, which is computed, following the OCC model, by the difference between potential and threshold.

We are aware that the proposed formalization is just a first step in an extension of the logical BDI framework for incorporating emotions. An emotional BDI model should address other important dynamic processes between emotions and the mental states of desires, intentions and beliefs in the BDI model. As the BDI is a practical reasoning architecture, *i.e.*, reasoning towards action (Wooldridge, 1999), it is important to discuss how the use of emotions can help agents to choose the most rational action to be performed, and how the emotions can improve the way that agents reason or decide or act. These are open questions that we intend to address in a future work. However, we believe that the logic formalization of the appraisal modeling, presented in this paper, is an important initial point for the specification of an emotional BDI architecture, since the appraisal evaluation explains the origin of an emotion and also differentiates them (Scherer, 1999). Another important future work is the formalization of the other sixteen emotions of the OCC model.

Acknowledgments

References

Adam, C., Herzig, A., & Longin, D. (2009). A logical formalization of the OCC theory of emotions. *Synthese*, *168*(2), 201–248.

Bagozzi, R. P., Dholakia, U. M., & Basuroy, S. (2003). How effortful decisions get enacted: the motivating role of decision processes, desires, and anticipated emotions. *Journal of Behavioral Decision Making*, *16*(4), 273–295.

Bratman, M. (1990). What is intention?. In Cohen, P. R., Morgan, J. L., & Pollack, M. E. (Eds.), *Intentions in Communications*, Bradford books, pp. 15–31. MIT Press, Cambridge.

Cassell, J. (2000). *Embodied conversational agents*. MIT Press, Cambridge, MA, USA.

Cohen, P. R., & Levesque, H. (1990). Intention Is Choice with Commitment. *Artificial Intelligence*, *42*(2-3), 213–261.

Dennett, D. C. (1971). Intentional Systems. *The Journal of Philosophy*, *68*(4), 87–106.

Dennett, D. C. (1997). True Believers: The Intentional Strategy and Why It Works. *Mind Design*, pp. 57–79.

Dias, J. a., & Paiva, A. (2013). I want to be your friend: establishing relations with emotionally intelligent agents. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, AAMAS '13, pp. 777–784, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Fagin, R., & Halpern, J. Y. (1994). Reasoning about knowledge and probability. *J. ACM*, *41*(2), 340–367.

Gebhard, P. (2005). Alma: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pp. 29–36, New York, NY, USA. ACM.

Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Texts in Statistics. Springer Science+Business Media, LLC.

Gluz, J. C., & Jaques, P. A. (2014). A Probabilistic Implementation of Emotional BDI Agents. In *6th International Conference on Agents and Artificial Intelligence (ICAART)*. INSTICC.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*(4), 269–306.

Gratch, J., & Marsella, S. (2005). Evaluating a computational model of emotion. *Autonomous Agents and Multi-Agent Systems*, *11*(1), 23–43.

Halmos, P. (1974). *Measure Theory*. No. v. 9 in Graduate Texts in Mathematics. Springer.

Halpern, J. Y. (2005). *Reasoning About Uncertainty*. Mit Press.

Halpern, J. Y. (1990). An analysis of first-order logics of probability. *Artificial Intelligence*, *46*(3).

Hodkinson, I., & Reynolds, M. (2007). Temporal Logic. In Blackburn, P., van Benthem, J., & Wolter, F. (Eds.), *Handbook of Modal Logic*, Studies in Logic and Practical Reasoning, pp. 550–620. Elsevier Science.

Hughes, G. E., & Cresswell, J. (2012). *A New Introduction to Modal Logic*. Taylor & Francis.

Isen, A. M., & Patrick, R. (1983). The effect of positive feelings on risk taking: When the chips are down. *Organizational Behavior and Human Performance*, *31*(2), 194–202.

Jaques, P. A., Vicari, R., Pesty, S., & Martin, J.-C. (2011). Evaluating a Cognitive-Based Affective Student Model. In D'Mello, S. K., Graesser, A. C., Schuller, B., & Martin, J.-C. (Eds.), *International Conference on Affective Computing and Intelligent Interaction (ACII)*, Vol. 6974 of *Lecture Notes in Computer Science*, pp. 599–608. Springer.

Jiang, H., Vidal, J., & Huhns, M. N. (2007). *International Conference On Autonomous Agents*. ACM, New York.

Meyer, J.-J., & Veltman, F. (2007). Intelligent Agents and Common Sense Reasoning. In Blackburn, P., Benthem, J. V., & Wolter, F. (Eds.), *The Handbook of Modal Logic*, pp. 550–620. Elsevier.

Monk, J. D. (1976). *Mathematical logic*. Springher-Verlag, New York.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119–124.

Ortony, A., Clore, G. L., & Collins, A. (1990). *The Cognitive Structure of Emotions*. Cambridge University Press.

Picard, R. W. (2000). *Affective Computing*. University Press Group Limited.

Raghunathan, R., & Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. *Organizational Behavior and Human Decision Processes*, *79*(1), 56–77.

Rao, A. S., & Georgeff, M. (1995). BDI Agents: from Theory to Practice. Tech. rep. Technical Note 56, Melbourne, Australia.

Rao, A. S. (2009). AgentSpeak(L): BDI Agents speak out in a logical computable language. *Seventh European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pp. 1–14.

Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence. Pearson Education/Prentice Hall.

Sadek, M. D. (1992). A Study in the Logic of Intention. In Nebel, B., Rich, C., & Swartout, W. (Eds.), *KR92 Principles of Knowledge Representation and Reasoning Proceedings of the Third International Conference*, pp. 462–473. Morgan Kaufmann.

Scherer, K. R. (2000). Psychological models of emotion. In Borod, J. (Ed.), *The neuropsychology of emotion*, Vol. 137 of *The neuropsychology of emotion*, chap. 6, pp. 137–162. Oxford University Press.

Scherer, K. R. (1999). Appraisal theory. In Dalgleish, T., & Power, M. (Eds.), *Handbook of Cognition and Emotion*, Vol. 19, chap. 30, pp. 637–663. John Wiley & Sons Ltd.

Searle, J. R. (1998). *Mind, language, and society : philosophy in the real world.* Basic Books, New York, NY.

Signoretti, A., Feitosa, A., Campos, A. M., Canuto, A. M., Xavier-Junior, J. C., & Fialho, S. V. (2011). Using an affective attention focus for improving the reasoning process and behavior of intelligent agents. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 02*, WI-IAT '11, pp. 97–100, Washington, DC, USA. IEEE Computer Society.

Sindlar, M. P., Dastani, M. M., & Meyer, J.-j. C. (2009). BDI-Based Development of Virtual Characters with a Theory of Mind. In Ruttkay, Z. (Ed.), *Intelligent Virtual Agents*, pp. 34–41, Berlin. Springer-Verlag.

Steunebrink, B. R. (2010). The Logical structure of emotions. *SIKS Dissertation series*, *2010*.

Steunebrink, B. R., Dastani, M., & Meyer, J.-J. (2008). A formal model of emotions: Integrating qualitative and quantitative aspects. In *European Conference on Artificial Intelligence (ECAI)*, Vol. 178, p. 256, Patras, Greece. IOS Press.

Steunebrink, B. R., Dastani, M., & Meyer, J.-J. (2011). A formal model of emotion triggers: an approach for BDI agents. *Synthese*, *185*(S1), 83–129.

van der Hoek, W., & Wooldridge, M. (2012). Logics for multiagent systems. *AI Magazine*, *33*, 92–105.

Van Dyke Parunak, H., Bisson, R., Brueckner, S., Matthews, R., & Sauter, J. (2006). A model of emotions for situated agents. *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems AAMAS 06*, *2006*, 993.

Warner, S. (1990). *Modern Algebra*. Dover Books on Mathematics. DOVER PUBN Incorporated.

Wooldridge, M. (2009). *An Introduction to MultiAgent Systems.* John Wiley & Sons.

Wooldridge, M. (1999). Intelligent Agents. In Weiss, G. (Ed.), *Multiagent systems*, pp. 27–77. MIT Press, Cambridge, MA, USA.