Emotion in Intelligent Virtual Agents: the Flow Model of Emotion

Luís Morgado[1,2] and Graça Gaspar[2]

[1] Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro, 1949-014 Lisboa, Portugal
lm@isel.pt
[2] Faculdade de Ciências da Universidade de Lisboa
Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
gg@di.fc.ul.pt

Abstract. Different models have been proposed to support emotion in artificial agents. However, a general framework to support the implementation of emotional agents of different kinds and levels of complexity is still not well defined. In this paper we present a model that is independent of specific physiological or psychological details, despite being inspired by biological processes, defining an emotional structure that can be objectively implemented and evaluated, and which can be used to integrate and extend other agent models, like deliberative models. Concrete results are presented to illustrate the model adequacy for agent emotional characterization.

1 Introduction

Two main approaches have characterized the development of emotion models for intelligent agents, physiologically inspired models (e.g. [2, 8]) and appraisal theories inspired models (e.g. [4, 1]). These models have allowed the definition of emotion like characteristics and behavior, however they also have some drawbacks. Physiologically inspired models are based on specific mechanisms of biological organisms, some of them still not well known, like hormonal mechanisms, resulting in highly specific implementations limited to relatively simple agents and contexts. On the other hand, in appraisal theories based models two main problems arise: (i) because the evaluation and interpretation of events is based on appraisal dimensions that are defined in subjective psychological terms, it is difficult to define objective operational criteria for appraisal, resulting in empiric and also specific implementations; (ii) appraisal theories emphasize the structural aspects of emotion elicitation, but don't say much about the underlying processes, as Scherer refers "… most research as been firmly wedded to the notion of emotional state and its assessment via verbal labels" [13]. Besides these problems, the aspects of adaptation and learning, which are directly related to emotion as recent studies show, are not specifically addressed.

To effectively support emotion in intelligent agents it is necessary to address these problems and to define emotional models that can serve as a general framework for the implementation of agents of different kinds and levels of complexity. It is also important that these models be able to integrate and extend other agent models, like deliberative models (e.g. BDI models) to support emotional aspects. This is one of the main aims of the work that we are conducting.

In this paper we present the motivation and overall structure of such a model, which we call *Flow Model of Emotion*. The proposed model addresses these problems by: (i) defining an emotional structure that can be objectively implemented and evaluated; (ii) defining the dynamics underlying the emotional structure which is directly related to the issues of adaptation and learning.

The paper is organized as follows: in section 2, a model of emotional dispositions is presented; in section 3, we discuss how that model can be used for modeling emotional agents; in section 4, we describe how to emotionally characterize an agent; we conclude by presenting an example.

2 The Flow Model of Emotion

The classical approach to emotional appraisal suggests the existence of base appraisal dimensions that individuals use to evaluate the significance of events (e.g. [5, 7, 10, 12]). Different authors have proposed different sets of appraisal dimensions. However, there are two dimensions that are common to the main models (even if appearing under different names): goal conduciveness and coping potential as named by Scherer [12] or motive-consistency and control potential as named by Roseman *et al* [11], among others. From our point of view, these two dimensions are effectively base dimensions, while other proposed dimensions, such as pleasantness or norm compatibility are contextualized variations of this two base dimensions, namely in cognitive and social contexts. However, even these two dimensions serve only to classify an emotional state from an external perspective, but are not sufficient to explain the process through which emotional states are attained and emotional behavior is generated. In fact this is one of the above-mentioned main problems of appraisal theories of emotion.

Another dimension that is normally included as an appraisal dimension is novelty (or other variations as unexpectedness or surprise). From our point of view, this dimension reflects essentially the level of change, therefore it results from the dynamics of the emotional processes - it can be needed to characterize an emotional state from an external point of view, but it is not explicitly represented in terms of internal structures.

We propose that the two appraisal dimensions, goal conduciveness and coping potential, reflect the existence of base attributes that support the elicitation of emotional states and behavior. We call these base attributes *achievement potential* (P) and *achievement conductance* (C). P represents the potential of change that an agent is able to produce in the environment to achieve the intended state-of-affairs. C represents the degree of the environment's conduciveness or resistance to that change, which can also mean the degree of environment change that is conducive, or not, to the agent intended state-of-affairs. Therefore, achievement potential and achievement conductance represent two complementary aspects in the relation agent/environment, which some authors recognize as one of the main aspects underlying emotion [7].

These attributes are distinct from the appraisal dimensions in the sense that they are objective attributes that characterize the cognitive structure elements of an agent and not subjective psychological dimensions. This is possible due to the conception of an agent as an open system that maintains itself in a state far from equilibrium, yet keeping an internally stable overall structure. This kind of systems is known as dissipative structures [6].

In this kind of systems the achievement potential can be viewed as a force (*P*) and the achievement conductance as a transport property (*C*). The dynamics of the emotional process can therefore be characterized as a relation corresponding to a flow, called *achievement flow*, defined by:

$$F = C \, . \, P \tag{1}$$

That means the achievement flow is proportional to the achievement potential, where the proportionality factor is a transport property that reflects the degree of conduciveness or resistance to the achievement flow.

One fundamental notion is implicit in the proposed model, the notion of energy. In thermodynamics, energy is usually defined as the capacity to produce work [6]. In this context energy can be defined as the capacity of an agent to act or, in a wide sense, to produce change.

In an agent, this capacity to produce change (to achieve the intended state-of-affairs) exists in a potential form. The effective achievement of the intended state-of-affairs depends on the effective application of that potential, that is, on the flow of the accumulated energy. Therefore, potential and flow correspond to two complementary aspects of energy: (i) the potential corresponds to the accumulation of energy, independently of its application; (ii) the flow corresponds to the application of the capacity to produce change as movement of energy. This complementary and interchangeable nature of energy is mediated by the underlying structure of the agent and the environment, under the form of conductance, reflecting their nature and dynamics. From this dynamic relation between achievement potential and achievement conductance emerge behavioral forces constraining the behavior of the agent and determining the evolution of its emotional state.

2.1 Emotional Disposition

To understand the nature of those forces let us consider a function *hs* with arguments *P* and *C* that characterizes the hedonic state of an agent, representing the level of wellbeing of an agent in relation to the achievement of its motivators. The evolution of *hs* can be related to the variation of the achievement potential (*P*) and the achievement conductance (*C*), as follows:

$$dhs/dt \equiv f(\delta P, \delta C) \quad \text{where } \delta P = dP/dt \text{ and } \delta C = dC/dt \tag{2}$$

Independently of the nature of function *f*, it can be viewed as representing a motivating force leading to the change of the agent's hedonic state and the corresponding emotional state. Therefore function *f* represents a force underlying the emotion related behavior of an agent. Since we don't know the exact nature of that force we propose that it can be represented as a vectorial function *ED*, called *emotional disposition*, defined as follows:

$$ED \equiv <\delta P, \delta C> \qquad (3)$$

Functions *P* and *C* represent two independent and complementary aspects, therefore, at a given instant $t = \tau$, δP and δC can be represented as two orthogonal dimensions of variation of *ED*.

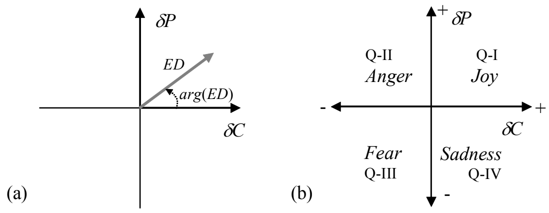

Fig. 1. Vector *ED* as a function of δP and δC (a); relation between *ED* quadrants and emotional quality tendency (b).

Since we consider *P* and *C* the main attributes upon which the elicitation of an emotional state is based, it is important to be able to relate their variation values with the emotional tendency that may be underlying them, as characterized from experimental evidence. In figure 1.b we identify four quadrants of evolution of *ED* as a function of δP and δC: Q–I ($\delta P > 0$ and $\delta C > 0$) “everything goes well”, directly related to joy; Q–II ($\delta P > 0$ and $\delta C < 0$) the agent has potential to handle the situation but there are “adversities”, typical of anger; Q–III ($\delta P < 0$ and $\delta C < 0$) the agent doesn’t have capacity to handle the “adversities”, typical of fear situations; Q–IV ($\delta P < 0$ and $\delta C > 0$) corresponds to a situation where despite no significant opposition to the achievement of the agent’s motivators, there is a negative variation of the agent’s potential; this kind of situation can be directly related to sadness, which according to experimental studies (e.g. [7]) often results from sudden losses (a sudden decreasing of *P* produces a strong negative variation of δP).

Each of the quadrants can be directly related to the emotional character of a specific emotion, which is represented by the orientation of the vector *ED*. The module of the vector *ED* corresponds to the intensity of the emotional disposition. That is:

$$Quality(ED) \equiv \arg(ED) \qquad (4)$$

$$Intensity(ED) \equiv |ED| \qquad (5)$$

The emotional tendency associated to each quadrant is only indicative of its main nature, since the quality of the emotional disposition is continuous. This is consistent with phenomenological well known emotion blends. In this way it is possible to account for the existence of basic emotional patterns without compromising the rich and continuous nature of emotional phenomena, as is the case in the classical composition of separate discrete emotions.

To conclude the presentation of the emotional disposition notion, we must consider the following aspects: (i) *emotional disposition* is defined as an action regulatory disposition or tendency, but it does not constitute in itself an emotion; (ii) *emotions* are considered emergent phenomena that result from the dynamics of agent/environment interaction. Their existence is ascribed by an external observer or by self-reflective processes of an agent, through the observation of specific behavioral patterns and states (either external or internal).

3 Modeling Emotional Agents: The Agent Flow Model

The emotional model just presented constitutes a conceptual framework for modeling emotional agents, however to support its concrete implementation an operational model is needed. We called that model *Agent Flow Model* [9]. In that model, two of the base notions of the flow model of emotion have architectural counterparts, namely, potentials and conductances. On the other side, the notion of flow corresponds to a transient aspect reflecting the dynamics of achievement.

Cognitive Elements. The architectural elements of the agent flow model have an inherent cognitive meaning corresponding to three basic cognitive notions: (i) *motivators*, elements that represent a motivating force for an agent to act, accumulated as achievement potential; (ii) *mediators*, elements that represent the media that supports the transformation of the motivating force into applied force to act, acting as flow conductors (conductances); (iii) *achievers*, composite elements that represent the commitment to apply the motivating potential through a specific mediator to generate the flow that realizes the applied force for an agent to act. These elements are the structural constitutive elements of an agent and can be arranged in multiple ways to produce different kinds of agents with different behaviors and levels of complexity.

Flow Networks. The main organizational patterns of the cognitive structure of an agent are networks of achievers through which the achievement flow will run producing action, both internal and external, to realize its motivators. Flow networks can be static or dynamic according to the type of the agent. In simple reactive agents, flow networks are static, determining a fixed set of agent behaviors. On the other hand, in deliberative agents, flow networks are dynamically generated by the cognitive processes, allowing a much more rich and versatile set of behaviors. More details on flow networks dynamics can be found in [9].

4 Agent Emotional Characterization

The emotional forces that result from the dynamics of the agent flow model can be made explicit for agent emotional characterization. To explicitly represent the emotional disposition we must define its scope. We call that scope a *cognitive unit*. A cognitive unit is composed by a set of motivators and the flow networks generated to achieve those motivators, and is characterized by an input flow (F_i), an achievement conductance (C) and an output achievement flow (F). The overall cognitive structure of an agent corresponds to a global cognitive unit representing the agent as a whole.

In a cognitive unit the instantaneous variation of the achievement potential (δP), depending on P, which is an internal characteristic of the cognitive unit, corresponds to the potential effectively applied at each instant, that is, to the achievement flow F. However, in concrete situations we must also take into account the fact that, in order to operate effectively, an agent must reach a minimum achievement flow level (F_{min}). This level is specific to an agent or a class of agents, and it may change during the agent activity. Therefore, δP corresponds to the effective achievement flow of an agent (F_e), leading to the following operational definition of emotional disposition:

$$F_e = F - F_{min} \tag{6}$$

$$ED_e \equiv <F_e , \delta C> \tag{7}$$

The explicit representation of the emotional disposition serves, essentially, for communication purposes, since its inherent emotional force underlies all the cognitive processes of the agent and its corresponding behavior. However, this communicative aspect is of particular importance in intelligent virtual agents, either to communicate between agents or to interact with humans. The following example serves to illustrate the descriptive adequacy and power of the proposed model.

4.1 Experimental framework

As a motivating example we will consider a personal assistant agent that supervises the financial investments on behalf of its human user. In this kind of situations emotion could play an important role both as motivator of the agent behavior as well as a highly effective tool for human-agent interaction.

The proposed example consists of a simulation where the agent experiments different degrees of achievement conductance, from all-conducive financial situations ($C = 1$), corresponding to high profits, to totally non-conducive financial situations ($C = 0$), corresponding to high losses. Two base scenarios were simulated, as illustrated in figure 2.a: (i) limited adversity; (ii) full adversity.

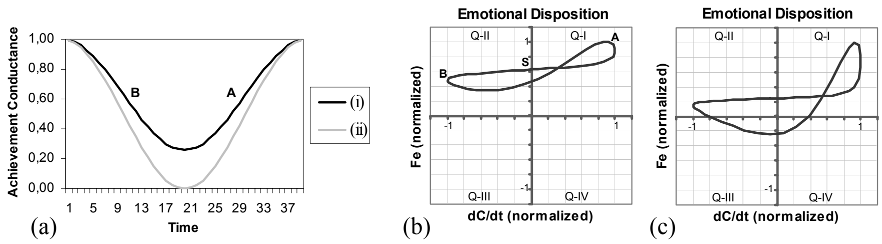

Fig. 2. Variation of achievement conductance with time (a). Resulting evolution of the emotional disposition (the time axis is omitted – successive *ED* vectors are projected in the emotional disposition plane – evolution direction is S-B-A) (b) and (c).

The motivational input flow that the agent generates to achieve its motivators was maintained constant. The minimum achievement flow (F_{min}) was fixed at 50% of the constant motivational flow.

4.2 Experimental results

At each time step the global emotional disposition of the agent was evaluated. The results for the two base scenarios (i) and (ii) are shown, respectively, in figures 2.b and 2.c. In experiment (i), as the conductance varies from high values to low values, the emotional disposition gains a clear anger tendency. When the conductance becomes high again, the agent turns to a joy tendency (satisfaction). However, if the conductance is further decreased, as is the case in experiment (ii), fear tendency appears and, before the agent turns to joy, the quadrant Q-IV is crossed, corresponding to a relief like emotional tendency.

Other interesting aspects can also be observed. For instance, considering the evolution of the emotional disposition intensity in experiment (i), shown in figure 3.a, it is possible to observe two distinct peaks that can be interpreted as reflecting an anger peak (B), resulting from the experienced adversities, followed by an elation/satisfaction peak (A) resulting from the end of the adversities.

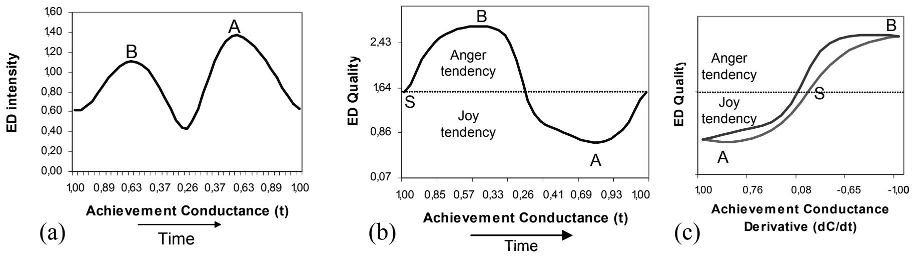

Fig. 3. Evolution of *ED* intensity as a function of the achievement conductance (a). Evolution of *ED* quality as a function of the achievement conductance (b), and as a function of the achievement conductance derivative (c).

Regarding the evolution of the emotional disposition quality it is interesting to note its nonlinear nature as shown in figure 3.b. In figure 3.c we can observe a hysteresis, typically characteristic of emotional responses. Increasing non-conduciveness leads to a point where an agent will in a dramatic fashion jump to a clear tendency to anger, rather than continue to change emotional quality in a linear fashion and, in the return path, it will take a higher increase of conduciveness before a drop to the joy tendency will occur.

4.3 Discussion

These experiments show some interesting results which are comparable to biological emotional phenomena and not predicted by classical appraisal models [13], namely: sudden shifts in emotional disposition; dependence of the departure point; emotional decay; and a continuous nature allowing for complex blends or mixtures of emotions. On the other hand, the proposed emotional and agent models are not dependent on specific contexts or kinds of agents, supporting the extension of deliberative agent models, such as the BDI model, to include emotional aspects [9].

5 Conclusions and Future Work

In this work we presented the fundamentals of the emotional flow mechanisms and how these mechanisms can be used to support the implementation of emotional agents. The proposed flow model of emotion is based on three main premises: (i) emotion is related with two base attributes, *achievement potential* (P) and *achievement conductance* (C); (ii) the emotional disposition can be characterized as a force whose components are the instantaneous variations of P and C; (iii) the dynamics of the emotional process can be characterized as a flow reflecting the joint variation of P and C, which we called *achievement flow*.

Future research will aim at the characterization of the emotional state of the agent and further exploring the adaptive nature of the proposed models for modeling different types of agents.

Acknowledgments

This research has been partially supported by the program PRODEP III 5.3/2/2001 and the project MAGO2 (POSI/39351/SRI/2001) from FCT.

References

1. Bates, J.: The Role of Emotion in Believable Agents. Communications of the ACM 37(7) (1994) 122-125
2. Cañamero, L.: Designing Emotions for Activity Selection in Autonomous Agents. In R. Trappl, ed., Emotions in Humans and Artifacts, MIT Press (2000)
3. Damásio, A.: Descartes' Error: Emotion, Reason, and the Human Brain. Putnam PG (1994)
4. Elliot, C.: The Affective Reasoner: A Process Model of Emotions in a Multi-agent System. PhD thesis, Northwestern Univ., The Inst. for the Learning Sciences, Tech. Rep. 32 (1992)
5. Frijda N.: The Emotions. Cambridge Univ. Press (1986)
6. Kondepudi, D., Prigogine, I.: Modern Thermodynamics: From Heat Engines to Dissipative Structures. John Wiley & Sons (1998)
7. Lazarus, R.: Emotion and Adaptation. Oxford University Press (1991)
8. Lisetti, C.: Emotion Generation for Artificial Agents via a Hybrid Architecture. In Proc. of the Workshop on Emotion-Based Agent Architectures (1999)
9. Morgado, L. Gaspar, G.: Extending BDI with Emotional Dispositions for Adaptive Social Reasoning. In Proc. 2nd International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (2003)
10. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988)
11. Roseman, I., Antoniou A., Jose, P.: Appraisal Determinants of Discrete Emotions: Constructing a more accurate and comprehensive theory. Cognition and Emotion, 10(3) (1996) 24-277
12. Scherer, K.: Toward a dynamic theory of emotion: The component process model of affective states. Geneva Studies in Emotion and Communication, 1(1) (1987) 1-98
13. Scherer, K.: Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In Lewis, M. & Granic, I. (Eds.) Emotion, Development, and Self-Organization. Cambridge University Press (2000) 70-99