In D. Canamero and R Aylett (Eds): "Animating expressive characters for social interactions". John Benjamins, in press.

Dynamic Models of Mixed Emotion Activation

Valeria Carofiglio
Department of Informatics
University of Bari

Fiorella de Rosis
Department of Informatics
University of Bari

Roberto Grassano
Department of Informatics
University of Bari

1 Introduction

The ability to show affective behavior is recognized as one of the essential ingredients of believability in Embodied Animated Characters. If artificial agents cannot be built to 'feel' emotions (at least as far as internal bodily changes are concerned), they must at least be able to simulate this condition in their external appearance. Shallow or inner aspects of behavior may be influenced by emotions, such as face expressions, gesturing, movements but also decision making, argumentation style, instructional strategies (Sillince and Minors, 1991; Gmytrasiewicz and Lisetti, 2000; Staller and Petta, 2001)

The increasing number of recent studies that are aimed at introducing some form of affect in the computer attitude have a common origin in the seminal works of Carbonell (1980), Oatley and Johnson Laird (1987), Ortony, Clore and Collins (1988) and others. In particular, Ortony and colleagues suggested a categorization of emotions based on their activation factors that is acknowledged as the grounding theory from the large majority of application studies. The proposed formalisation methods were also inspired by the formalization of the emotion activation mechanism in terms of emotion-specific rules that was suggested by Ortony (1988 b) and owe identification of the variables that may influence this process to the deep analysis of Elliott and Siegle (1993).

The way complex phenomena like emotions may be modeled strongly depends on the envisaged application area. If, for instance, the application concerns a 2D embodied character that is sketchy in its appearance and is expected to show a limited range of behaviors (like in the case of MS-Agents), then a refined modeling approach is probably not needed: emotions with graded intensities and multiple emotions would be hard to show, in this case.

In other cases, however, more refined representations are needed and possible. This is the case of 3D characters that are designed to be highly realistic and whose face and bodies may be manipulated so as to (ideally) show the large variety of expressions that are displayed by humans. Another situation of high realism occurs when the goal of realism and believability concerns the social behavior of characters rather than (or in addition to) their external apparence: for instance, when they are expected to undertake 'natural conversations' with the user. In the latter case, some knowledge of the *reasons why* an emotional state was activated (in the agent and in the user) is useful to achieve consistency and naturalness in the behavior, in a given time instant and throughout time. Models have then to be refined enough to represent the cognitive and social aspects of emotion activation and disappearing. For instance, which beliefs and goals govern activation of every emotion, how differently agents react to similar situations according to their personality and to the context in which the event occurs, which variables affect emotion intensities and, finally, how emotions may mix up and vary in time. In these case, a simple representation of (event, emotion) combinations is not sufficient and a finer grain size of knowledge is needed. In addition, time evolution of the affective state has to be represented.

In this paper, we discuss this problem in general terms, starting from Picard's 'Marathon example' (Picard, 1997, p. 171). We then propose a method to represent cognitive models of emotion activation and disappearing which deals with uncertainty and value scales.

2 Multiple Emotions.

In the 'Pure vs Mixed Emotions' Section of her book on 'Affective Computing', Rosalind Picard introduces the following example:

"After winning a Marathon, a professional runner described *feeling tremendously* ***happy*** *for winning the race,* ***surprised*** *because she believed she would not win, somewhat* ***sad*** *that the race was over and a bit* ***fearful*** *because during the race she had acute abdominal pain* (Picard, 1997, p. 171). A runner's friend, when assisting to that Marathon, was probably *"****happy for her*** *because she won the race, although a bit* ***envious*** *for not being able to participate to it and* ***sorry for*** *seing her so tired, at the end.*"

How is it that these two persons referred to feel this different mixture of emotions? Clearly, the main source of difference is due to the different structure of beliefs and goals of their minds. In the runner, the *intensity* of ***fear*** during the race was probably related, at the same time, to the *importance* she assigned to her goal of winning it and to variations in the *probability* of achieving this goal, that she dynamically revised during the race. The importance of this goal also affected the intensity of ***happiness*** (or ***satisfaction***) of achieving it, while ***surprise*** was probably due to a difference between the likelihood she attached to achieving the goal at the beginning of the race and the final result. The ***sadness*** that the event was over might be a mixed emotion in its turn, some combination of ***nostalgia*** for a pleasant past event and ***hope*** to be again in a similar situation, in the future. The mixing of emotions in the runner's friend was probably due to a mixing of goals of approximately equal weight: ***happy-for*** is due to his desire of achieving "the good of his friend" (conditioned to her winning the race) ***sorry-for*** to his desire of "preserving her from bad" (illness, in this case) and ***envy*** to his desire of "dominating her", in a way.

So: differences between the two persons, in the example, are due to differences in the beliefs, the goals they want to achieve, the weights they assign to achieving them and the structure of links between beliefs and goals. Variations of these measures with time seem to govern cognitively-generated emotions. Picard evokes the *generative mechanism* as the key factor for distinguishing between emotions that may *coexist* (by mixing according to a 'tub of water' metaphor) and emotions that *switch* from each other in time (by mixing according to the "microwave oven" metaphor). She suggests that co-existence may be due, first of all, to the differences in these *generative mechanisms*. But they may be due, as well, to differences in the decay speed between emotions that were generated by the same mechanism in two distinct time instants: for instance, "primary" emotions, like *fear,* and cognitively-generated ones, like *anticipation*.

To represent the two ways emotions may mix up, the modeling formalism adopted should therefore enable representing their generative mechanism, the intensity with which they are triggered and the way this intensity decays with time. We claim that *dynamic belief networks* (DBNs) are an appropriate formalism to achieve the mentioned goals. We show, in adition, how they enable representing the dynamic arousal, evolution and disappearing of emotions and the way these phenomena are influenced by personality factors and social context. We illustrated in detail the logics behind this modeling method in another paper, in which we discussed the advantages it offers in driving the affective behavior of an Embodied Conversational Agent that we are building in the scope of the European Project "Magicster"[i] (de Rosis et al, in press).

In this contribution, we summarise the main features of our modeling method, to focus our discussion on how the two metaphors of mixing emotions proposed by Picard may be represented with this formalism. We restrict our analysis to goal-based emotions, in Ortony's classification (Ortony et al, 1988). We briefly describe, in Section 3, what are DBNs and how they may be employed as a monitoring tool to represent emotion triggering. We then focus our analysis on how differences in the cognitive generative mechanism of emotions may be represented with DBNs, with the aim of establishing a correspondence between the differences that might occur in these mechanisms and the two mixing metaphors proposed by Picard. We finally contrast our method with some of the alternative approaches that have been proposed recently.

3. Emotion triggering with DBNs.

As we anticipated, our departure point is that emotions are activated by the belief that a particular important goal may be achieved or threatened. So, our simulation is focused on the *change* in the belief about the achievement (or threatening) of goals of an agent A, over time. In our monitoring system, the cognitive state of A is modeled at the time instants {T, T+1, T+2, …}. Events occurred during the time interval (T,T+1) are observed, to construct a probabilistic model of the new state and reason about emotions that might be triggered by these events. A well known formalism for representing dynamic phenomena in conditions of uncertainty is that of DBNs.

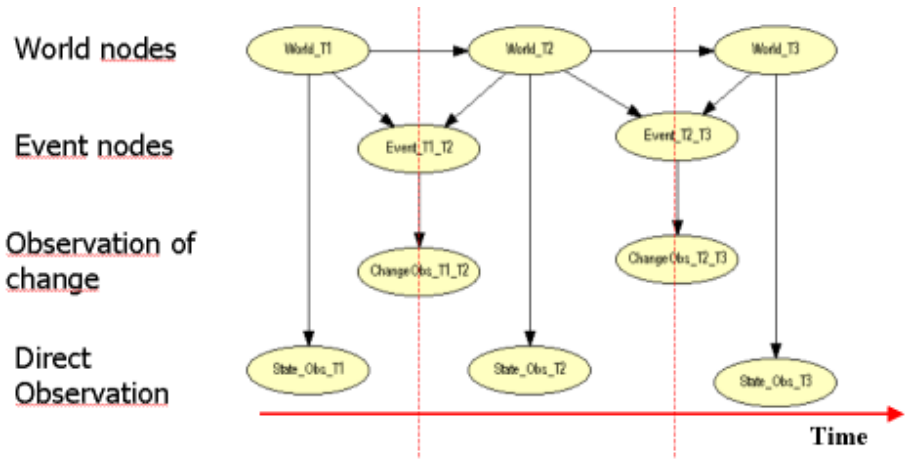


Figure 1: *Dynamic belief networks in monitoring systems (from Nicholson and Brady, 1994)*

As shown in figure 1, DBNs are based on the idea that time is divided into time slices, each representing the state of the modelled world at a

particular instant Ti. This state is described by means of a static belief network: World_Ti in the figure, with its observable state variables State_Obs_Ti. When DBNs are employed for monitoring purposes, two consecutive time slices are linked by arrows between the domain variables that have to be monitored. When something changes in the world, an event Event_Ti_Ti+1 occurs, that is observed through the variables in Change_Obs_Ti_Ti+1. The network is then extended for an additional time slice Ti+1. As a consequence, its structure and the probabilities of its nodes usually change (Pearl, 2000). To avoid explosion in the complexity of the network (and therefore in the uncertainty propagation algorithm), pruning of time slices and of network parts is performed after a new observation is added to the model, with a mechanism of roll-up (Nicholson and Brady, 1994).

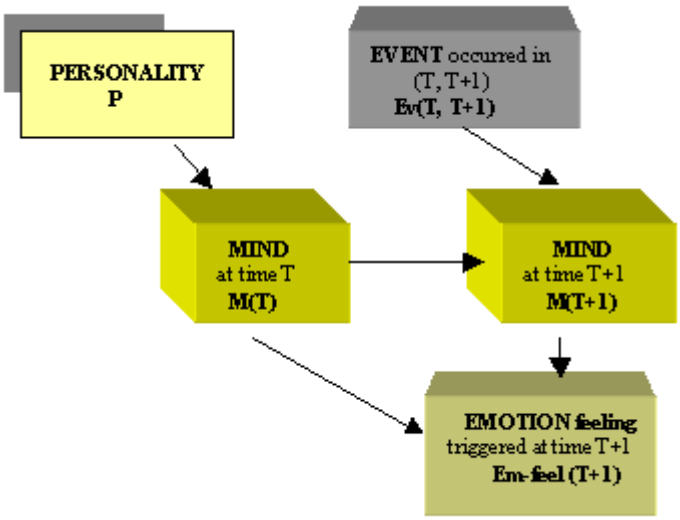


Figure 2: *Outline of our Emotion Monitoring System*

Figure 2 shows the general structure of our model of emotion activation, that includes the following static components:

1. M(T) represents the agent's Mind at time T, with its beliefs about the world and its goals;
2. Ev(T,T+1) represents the event occurred in the time interval (T, T+1), with its causes and consequences;
3. M(T+1) represents the agent's Mind at time T+1;
4. Em-feel(T+1) represents the fact that a particular emotion is activated, in the agent, at time T+1.

M(T+1) depends on M(T) and the event occurred in the interval (T,T+1). The feeling of emotions depends on both M(T) and M(T+1). We calculate the intensity of emotions as a function of two parameters: (1) the *uncertainty* in the agent's beliefs about the world and, in particular, about the possibility that some important goal is achieved or threatened, and (2) the *utility* assigned to achievement of this goal. More in depth, if:

a. A denotes the agent; G_i a high-level goal and Ach-G_i the achievement of this goal;
b. Bel A Ach-G_i denotes A's belief that the goal G_i will be achieved;
c. P(Bel A Ach-G_i) and P*(Bel A Ach-G_i) denote, respectively, the probabilities that A attaches to this belief, before and after the event Ev occurred;
d. W_A (Ach-G_i) denotes the weight that A attaches to achieving G_i,

then, according to the utility theory (Pearl, 1990), the *variation of intensity in the emotion* (ΔIe) may be calculated as follows:

$$\Delta Ie=[P^*(\text{Bel A Ach-}G_i)-P(\text{Bel A Ach-}G_i)] * W_A(\text{Ach-}G_i)$$

In other words, ΔIe is the product of the change in the probability that G_i will be achieved, times the weight of this goal. In negatively-valenced emotions (such as fear, sorry for etc), the probability that a goal G_i will be threatened (Thr-G_i) comes into play, instead of its achievement.

4. Goal-Based Emotion Categories

"Fortune-of-others" emotions (*sorry-for, happy-for, envy and gloating*) may be represented as points in the two-dimensional space ("desirability of the event", "empathic attitude"). Happy-for and envy apply to "desirable" events while "sorry-for" and "gloating" apply to "undesirable" ones; happy-for and sorry-for are driven by an empathic attitude, while gloating and envy are driven by a non empathic (or even contrasting) one.

Figure 3 shows the dynamic belief network that models how *happy-for* and *envy* may be activated in an agent A who assists to her friend U running the Marathon. This model shows that *happy-for* is triggered after believing that U will win the run or at least will come to the end of the race with a good time. If this event occurs, the probability of the belief that the high-level goal of 'desiring the good of others' will be achieved (Ach-GoodOf U) increases. The intensity of this emotion depends on the variation of this probability, that is produced when evidence about this 'desirable' event is propagated in the network. It depends, as well, on the weight the agent attaches to achieving that goal; this weight is, in its turn, a function of the agent's personality. It is high for *altruistic* people, low for *egoistic* ones. The emotion of *sorry for* is triggered by the same variables as *happy-for*.

However, some of these variables have, in this case, opposite values than in the case of *happy-for*, because they are related to *undesirable* occurring events: again, in our example, when the runner seems to be suffering. The high-level goal involved in this case is that of "preserving others from bad" (Ach-PresBad U). *Gloating* is related to envy by a similar triggering mechanism: the high-level goal involved is, in this case, that of "desiring the bad of others" and "preserving self from bad", at the same time.

Which "fortune of others" emotions may mix up and how? If a cognitive model of emotion activation is taken (as in our case), a correspondence may be established between cognitive generation of emotions and the set of beliefs and goals that influence the variation of the probability of achievement (or threatening) of the goals that govern their activation.

Again, let us consider the BN for activation of envy and happy-for that is shown in figure 3; let us consider, in particular, the emotion of happy-for. As stated, the involved goal is, in this case, 'achieving the good of others' (in particular, of U): happy-for is triggered by the belief that achieving this goal (Bel A (Ach-GoodOf U) increases over a given threshold. The root nodes of this subnetwork, that may influence variation of this probability and are directly influenced by the considered event, are the following:

Bel A (Has U x) :*"the agent believes that U has something x"*
Bel A (Desirable x) *"that is desirable"*.

If envy is considered instead, the threatened goal is 'dominating others' and the root nodes are the same as for happy-for, plus the additional belief:

Bel A not (Has A x) that *"the agent itself has not x"*.

The cognitive generation mechanism may therefore be directly represented, in BNs, in *"the root nodes of the subnets affecting goal achievement or threatening, in the Agent's mental state"*. In this representation, two emotions can coexist iff and only if their cognitive generation mechanisms are compatible: that is, if all the root nodes in their activation subnets take compatible values. According to this model, *happy-for* and *envy* are examples of potentially coexisting emotions. As we saw in figure 3, *the set of root nodes that their generation subnets have in common take compatible values when evidence about some observed event is propagated in the net*: in particular, the nodes Bel A (Desirable x) and Bel A (Has A x) must both be true. So, agents who are moderately altruistic and moderately dominant may be moderately envious and moderately happy-for at the same time, when they come to know that a desirable event occurred to a friend. For similar reasons, *sorry-for* and *gloating* may coexist, in particular circumstances (we don't show the DNB for this case).

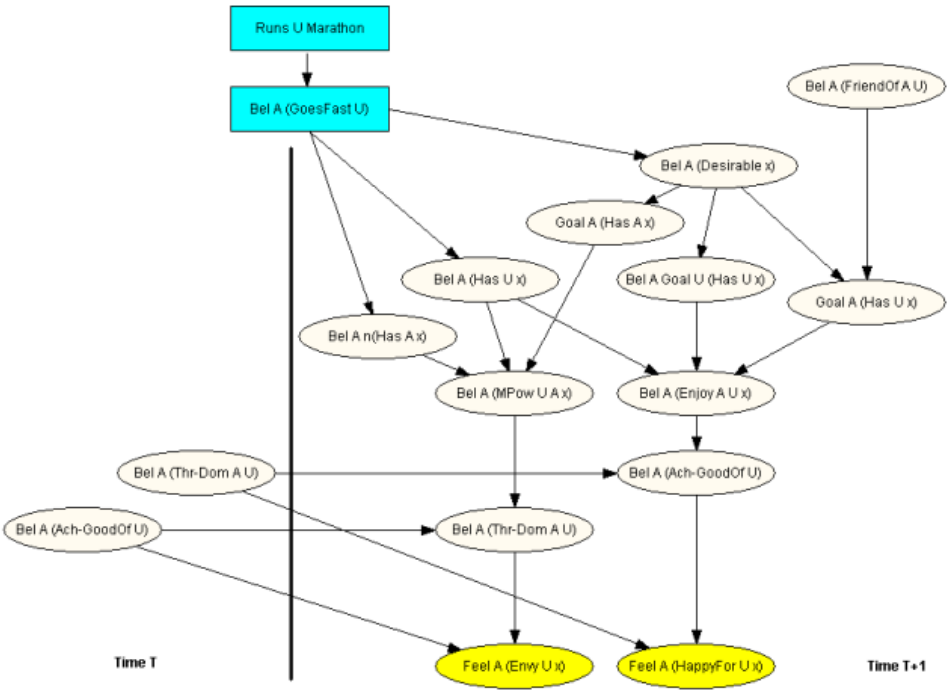


Figure 3: *The DBN for envy and happy-for*

Does the same mixing metaphor apply to other emotion combinations, in the category of 'fortune of others'? Probably not. As we said, they are triggered by a different value of desirability of the occurring event and, therefore, by *incompatible values of some common root nodes* in their generation subnets. Therefore, they cannot coexist. However, it might happen that the same event *activates*, in two different time instants, *different kinds of beliefs*. Persons may see positive and negative consequences of the same event in different times. In our example (see figure 4) the agent may notice that *"U is going fast"* in the interval (T, T+1), hence feeling *happy-for* (and, as we said, maybe also a bit of *envy*). It may subsequently notice (in T+1, T+2) that *"U seems to be tired"*: if the second event occurs before the effect of the previous one disappeared, the emotional state of sorry-for will co-occur with the previous one, according to a 'tub of water' effect. Otherwise, the Agent will switch between the two (positive and negative) emotional states, according to a 'microwave oven' effect.

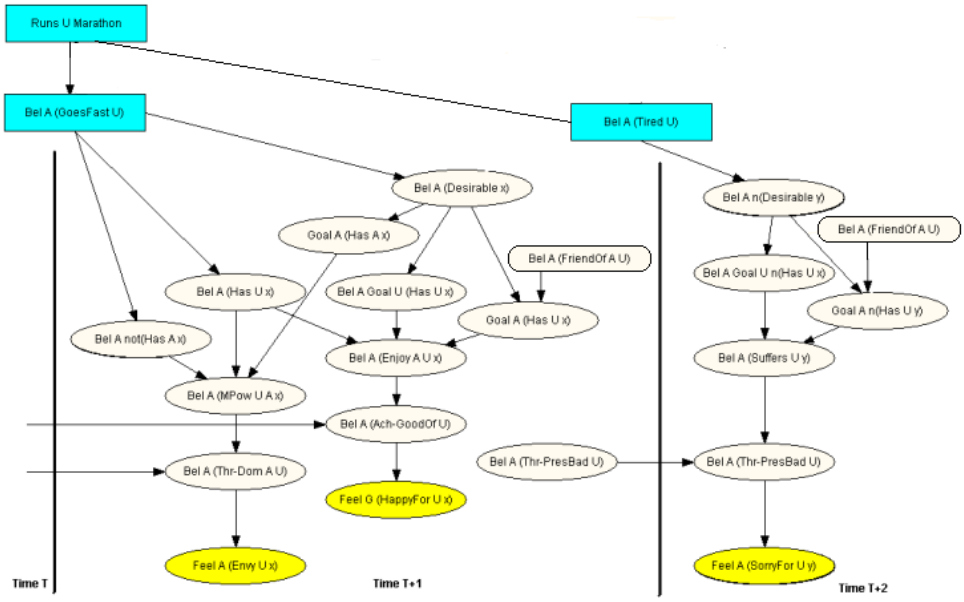


Figure 4: T*he DBN for happy-for and sorry-for*

More or less fast switching in time may occur from ***Prospect-based*** emotions (*fear, hope*) to ***Well being*** (*distress, joy*) or ***Confirmation*** emotions (*disappointment, relief*). Belief about achievement or threatening of high-level goals ("Desiring the Good of Self" or "Preserving Self from Bad") are involved in these cases. Switching from one emotion to another is due to a change in the probability of the belief that a (desirable or undesirable) event will occur, is occurring or occurred. This change may be due to observation of different evidences originating from this event at different times. In the Marathon example, switching from ***fear*** to ***joy*** to ***relief*** is closely related to the probability with which the runner believes she will win the race; this probability changes from time to time, according to the progress of the race. A "microwave oven" metaphor may therefore be applied to mixing of emotions originating from a change of value of some belief at different time instants, in the three categories. The same happens for emotions within the Prospect-based, the Well-being or the Confirmation category, which cannot coexist but between which the Agent may switch. For instance, the Marathon runner was, in different time instants, ***hopeful*** to win the race and ***fearful*** about its possible consequences on her health. According to similar considerations, one may conjecture about mixing of emotions belonging to different categories: for instance, ***happy*** and ***happy-for*** may coexist, as well as ***sorry-for*** and ***relief***, … and so on.

5. A Tool for simulating mixed emotions

The best way to validate representativeness of a model is to check its behavior in some 'typical' application domain. To test our emotion activation modeling method, we built a testbed with a graphical interface (that we called Mind-Testbed) that we applied to domains of various complexities. A model may be built, tested and refined, with Mind-testbed, through a iterative process:

1. ***In the building step***, the following components are set up:

- *the agent's mental* state, as a belief network;
- for every application domain, *the events that may occur* in this domain, again as belief networks;
  the set of *personalities* the agent may take: every personality is a psychologically plausible combination of *traits;* every trait corresponds to assigning a *weight* to a goal-node in the agent's mind. Figure 5 shows an example of personality: a 'other-centered' agent gives a high weight to the goals of achieving the good of others and preserving others from bad. If the agent is also optimistic, it gives a high weight, as well, to the goal of achieving the future good of self;

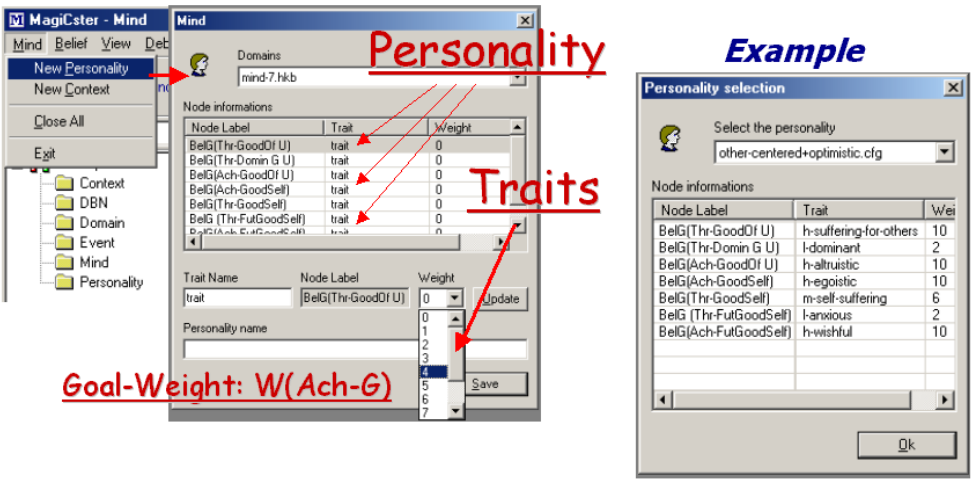
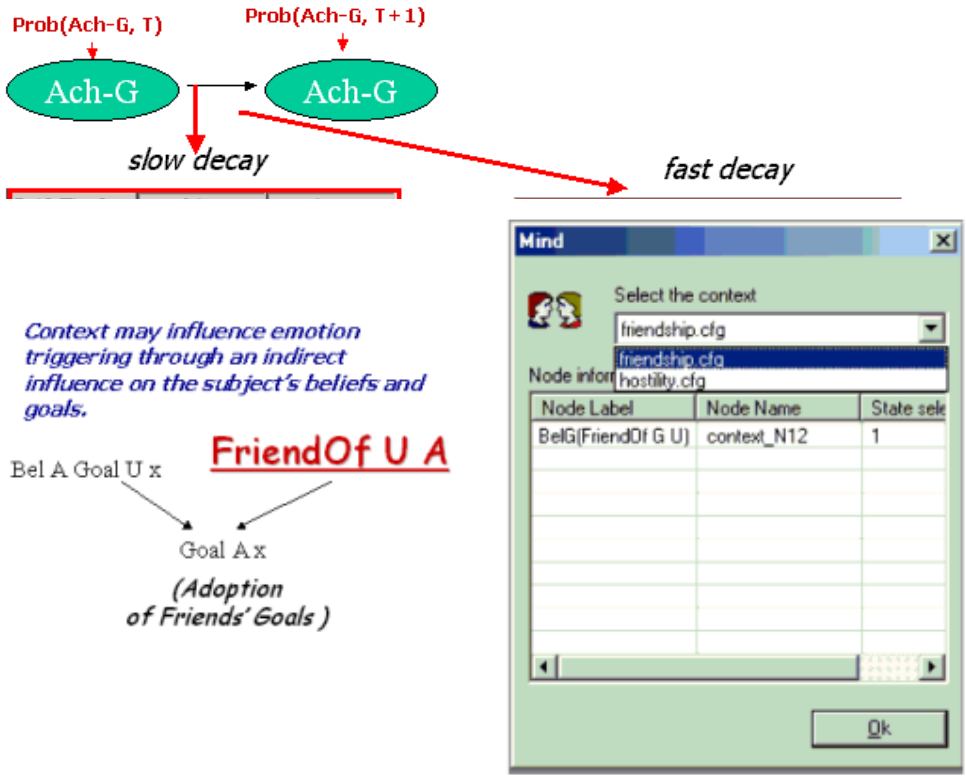


Figure 5*: Defining personality traits*

- the set of *contexts* in which interaction with the environment may take place; a context corresponds to assigning a value to one or more *context-nodes* in the Agent's mind. Figure 6 shows an example in which the 'social' context is set up, to be employed as shown (for instance) in figures 3 and 4 for the models *of envy, happy-for and sorry-for*. A



friendship context involves adopting the interlocutor's goal as far as possible.

Figure 6*: Defining contexts*

- a model of the *relationship between emotions and goals* in the agent's mind: in these models, the way every emotion *decays* is mirrored in the conditional probability table associated with the link between goal nodes at time T and T+1. The highest the probabilities assigned to the (true, true) and the (false, false) combinations, the slower is the emotion decay (as shown in figure 7). Decay time may be defined differently for the different emotions and may be varied according to another trait of the agent's personality: how persistent it is in its mood.

Figure 7: *Setting emotion decay*

2. ***In the testing step,*** a graphical interface enables selecting a *personality* for the Agent, a *context* in which interaction will take place and a *domain* (figure 8). A *threshold* for the intensity is settled for the emotion activation. A sequence of events is then fired and the way the agent's emotional state changes with time as a consequence of these events is monitored (in tabular and in graphical form) in the bottom frame.

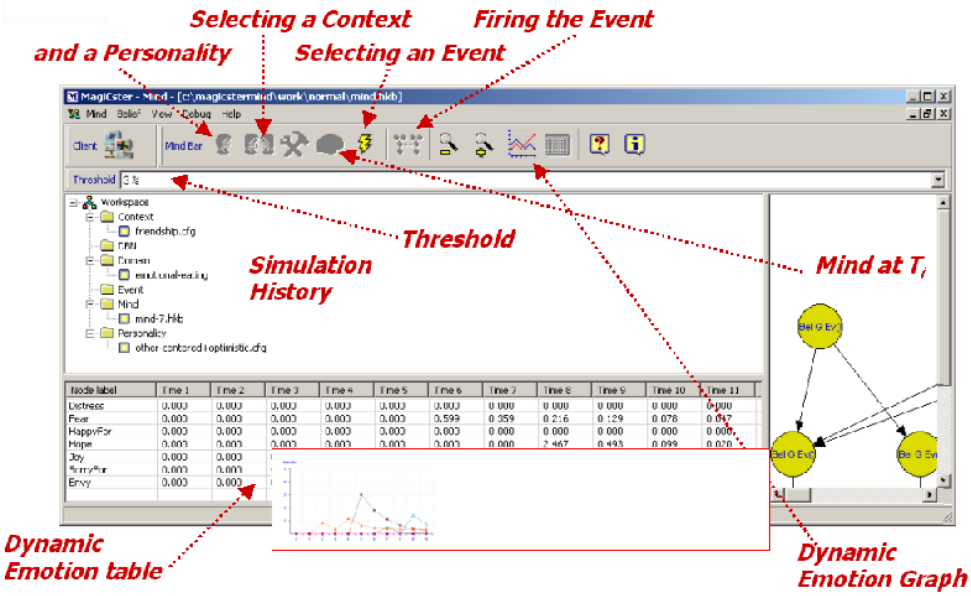


Figure 8: *The interface in the testing phase*

6. An Example

Let us consider again Picard's marathon example, to imagine the following dialog between the agent (A) and his friend Uta (U):

```
A0:Hi Uta. How are you?
U0: I'm OK. I'm planning to run the Boston Marathon.
A1: Yes I know. Are you in good training shape?
U1: My coach says so.
A2: Good, I'm happy for you.
U2: And what about you? Are you coming to Boston too?
A3: No, unfortunately I can't. Do you feel OK?
U3: I'm not at my best. I got injured on my last run.
A4: Oh! I'm sorry for you!
U4: Nothing serious.
A5: And now?
U5: Now I'm fairly OK. What about a small run together?
A6: Sure, with pleasure!
```

The agent talks at times T0, T1, T2, … The interlocutor's moves, in the intervals between two time instants, are seen as events that may trigger an emotional state in the agent's mind. In the example, the agent feels a mixture of happy-for and envy at time T2 (move A2), after Uta declares that she will run the next Boston Marathon (a desiderable event). At time T4 (move A4), after Uta declares not to be in her best conditions, the agent feels sorry-for, which may overlap to some still lasting happy-for and envy.

Finally, at time T6 (move A6), joy is triggered from the desirable event (to A) of going to run with Uta; this emotion overlaps with a very low (due to the decay) happy-for and sorry-for. The emotions felt by the agent are represented, first of all, in the content and the style of its dialog moves. At move A2:*" Good, I'm happy for you!"* manifests happy-for and a bit of envy. At move A3, *"Unfortunately"*, manifests a little distress. At move A4: *"Oh, I'm sorry for you!"* manifests sorry-for. At move A6, *"With pleasure!"* manifests joy.

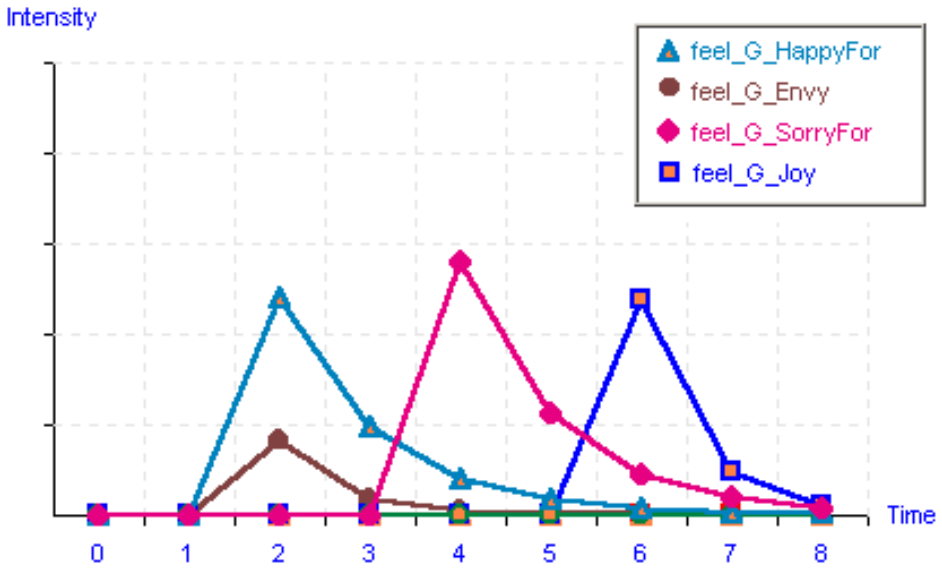


Figure 9: *Emotion activation in the Marathon dialog*

If the system takes the appearance of an embodied animated agent, it may show its emotional state through face expression, gesturing and so on, as we describe in (de Rosis et al, in press). In this case, multiple emotions will reflect into composite body expressions (Kaiser and Wehrle, this issue). The same model may be applied to recognize rather than generate emotional states. When the interlocutor shows an expression of some emotional state, the agent may try to guess the causes of this state. That is, it may try to infer the set of beliefs and goals that probably produced it, through some form of 'emotional plan recognition'. To do this, it will employ knowledge about the context, the interlocutor's personality and the events occurred in previous steps of the dialog.

7. Power of DBNs in Representing the Mixing of Emotions

In this paper, we described how DBNs enable representing the two kinds of situations that produce mixing of emotions:

1. *concurrent triggering* of emotions occurs when an event has *different consequences* whose beliefs co-occur in the Agent's mind and when these beliefs influence, in their turn, the achievement of different but *compatible* sets of beliefs and goals [ii];
2. s*witching* among different (and possibly contrasting) emotions occurs when the probability that the Agent attaches to a given belief changes with time or when new beliefs are activated in different time instants: these beliefs may be a consequence of the same or of different events.

We model time decay of emotions by introducing a link between two nodes representing the same goal at different time instants. As we said, the strength of this link may be varied, to simulate various forms of time decay. This enables us to formalise the case in *which coexistence is due to the fact that, after the same stimulus, emotions tend to take more or less time to disappear* and therefore may coexist for part of the time. It enables us, in addition, to consider the role played by another personality trait for our agent: its 'persistence in mood', that is a temperament characteristics that influences the tendency to stay in an emotional state after this state has been activated.

8. Related work

We are assisting to a flourishing of proposals on how to model emotion elicitation, expression and recognition: these methods all build upon psychological theories (Ortony et al, 1988; Elliott and Siegle, 1993; Oatley and Johnson-Laird, 1987, to cite only the main of them) and differ in the mathematical methods that are employed in formalising them. The main ingredients of emotion triggering models may be found in these theories. Oatley and Johnson-Laird suggest that "the functions of emotion modes are both to enable one priority to be exchanged for another in the system of multiple goals, and to maintain this priority until it is satisfied or abandoned" (Oatley and Johnson-Laird, 1986). Roseman establishes a relationship between 'appraisals', that are defined in terms of motivational and situational state, and emotion raising. He represents "problem type expectedness" and "control potential" in terms of beliefs (cited in Picard, 2000). Picard suggests that affective computational models should include a representation of "response decay" and of the influence of personality and temperament.

As far as the methods for formalising these theories are concerned, two main approaches are prevailing: the first of them follows the line traced by

Ortony in 1988, by proposing *mathematical functions* to combine a large number of (numerical) parameters in a measure of emotion intensity. In Affective Reasoner, Elliott lists the variables that may influence the intensity of an agent's affective state (Elliott and Siegle, 1993), without proposing any specific 'emotion-intensity calculation function'. The computational model of Affective Reasoner has been recently developed to formalise emotional reasoning in Emile (Gratch and Marsella, 2001). The main idea behind this system is to link emotion elicitation to appraisal of the state of plans in memory rather than directly to events. Emotion intensities are measured as a function of the goal importance and of the probability of goal attainment. This is evaluated, in its turn, by an analysis of the plans that could bring that goal about. In Prendinger et al (2002), this function is a logarithmic combination of the sum of exponentials of a number of variables: the intensity of *joy* depends on the desirability of the triggering event; the intensity of *happy-for* is a function of the Interlocutor's happiness and of the "degree of positive attitude" of the Agent towards the Interlocutor, and so on. If the number of variables is low, the function's value is only influenced by the "strongest" variable. Personality does not influence emotion elicitation but only the "filtering" process, in which it is decided whether the emotion should be expressed or suppressed. The main limit of this method is, in our view, in the combination of several heterogeneous measures in a unique function: when different scales are employed, the effect of their combination and the relative importance of each of them cannot be foreseen precisely. In addition, the range of variability of the function may go outside the established range of the variable that this function is presumed to measure, with the consequent need of "normalising" the obtained value or of "cutting" values out of this range.

The second approach to affective modeling is based on representing the cognitive aspects of emotion triggering. Belief Networks (BNs) seems to be a ideal formalism, in this case. BNs have first been applied to model expression and recognition of emotions from the Agent's external expression (voice, face, etc: Ball and Breese, 2000). DBNs have the additional advantage of enabling representation of the dynamic aspects of this phenomenon. They have been applied, recently, to model affective aspects of the User's mind in computer games (Conati, in press).

Our modeling method follows this second line of research. The main difference from the cited experiences is in the grain size of knowledge represented. Rather than representing directly the effect of events on the agent's mind, we build a fine-grained cognitive structure in which the appraisal of what happens in the environment is represented in terms of its effects on the agent's system of beliefs and goals. We only employ two measures (uncertainty of beliefs and utility of goals), while Elliott and Siegle propose a different measure for each variable. However, there is a

close relationship between the aspects of the phenomenon that are represented in the two cases. Elliott and Siegel's "importance of achieving a goal/not having a goal blocked" corresponds to the "weights" we attach to goal achievement or threatening. All variables associated with our "belief" nodes correspond to their "simulation event" variables (for instance, the "appealingness of a situation") and "stable relationship variables" (for instance, "how desirable an event is", "how friend is the agent with her interlocutor"). Rather than attaching to them an integer value in some established scale (as Elliott and Siegle proposes), we discretize these variables in a limited number of values and attach a probability distribution to each of them. In addition, by attaching conditional probability distributions to the network's links, we define a strength of these links in probabilistic terms.

The advantages of our formalism are in the possibility of representing the following aspects of emotion elicitation:

1. *temperament and personality*: how these factors affect the Agent's propensity to feel and to show emotions, by influencing the weights it assigns to achieving its goals, the threshold for emotion activation and the resistance to change its emotional state;
2. *social context:* how the Agent's relationship with the context influences emotion triggering by influencing its beliefs and goals;
3. *dynamics of the Agent's state:* how the Agent's affective state changes over time, either as a consequence of 'endogenous' or of 'exogenous' factors;
4. *response decay:* how emotion felt by the Agent decays, in absence of any new specific stimulus.

Of course, a model is always a shallow representation of a complex reality: the historical evolution of models in physics is a witness of the progressive advancement of knowledge in that domain. Refinement of a model requires, however, comparison with accurate experimental data in addition to a progress of mathematical techniques. This is true also for the domain of emotion modeling: in particular, refining parameters in BNs and the way they change with the context requires testing the model in well defined settings (which is probably of the same difficulty of experiments in modern physics). We plan to revise and refine the structure of our DBNs and their parameters through a systematic simulation of the kind of dialogs we want our artificial agent to make and through an evaluation of the 'believability' of these dialogs. More in general, if interest towards affective aspects of Human-Computer Interaction is not a transient reality, we will hopefully assist to a progressive refinement of these models, through an interleaving of theory formulation and experimental validation.

References

Ball, G. and Breese, J. 2000. "Emotion and personality in a conversational agent." In S. Prevost, J. Cassell, J. Sullivan and E. Churchill (Eds), *Embodied Conversational Agents*. pp. 189—219. MIT Press.

Carbonell, J.C. 1980. "Towards a process model of human personality traits." *Artificial Intelligence*, Vol. 15, pp. 49-74.

Conati, C. 2002 "Probabilistic assessment of user's emotions in educational games." *Applied Artificial Intelligence.* Special Issue on 'Merging cognition and affect in HCI'. F. de Rosis (Ed).

de Rosis, F. 2001. "Towards adaptation of interaction to affective factors." Preface to the Special Issue of *User Modeling ad User-Adapted* Interaction on 'User Modeling and Adaptation in Affective Computing'. 11, 4.

de Rosis, F. 2002. "Towards merging cognition and affect in HCI." Preface to the Special Issue of *Applied Artificial Intelligence* on 'Merging cognition and affect in HCI'.

de Rosis, F., Pelachaud, C. Poggi, I., Carofiglio, V. and De Carolis, C. (in press). "From Greta's mind to her face: modeling the dynamics of affective states in a Conversational Embodied Agent." *International Journal of Human-Computer Studies.* Special Issue on 'Applications of Affective Computing in Human-Computer Interaction'. E. Hudlicka and M NcNeese (Eds), in press.

Elliott, C. and Siegle, G. (1993). "Variables influencing the intensity of simulated affective states". *Proceedings of the AAAI Spring Symposium on Mental States '93, 58-67.*

Gmytrasiweicz, P. and Lisetti, C (2000): "Using decision theory to formalize emotions. In S.Parsons and P.Gmytrasiewicz (Eds), *Proceedings of the Second Workshop on Decision Theoretic and Game Theoretic Agents*, pp.39-47, Boston.

Gratch, J. and Marsella, S. (2001). "Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents". *Proceedings of the 5th International Conference on Autonomous Agents, Montreal, Canada, June.*

Hudlicka, E. (2002). "This time with feeling: integrated model of trait and state effects on cognition and behavior". *Applied Artificial Intelligence, Special Issue on 'Merging Cognition and Affect in HCI'*, F. de Rosis (Ed).

Kaiser, S. and Wehrle, T. (this issue). "Facial expressions in social interactions: beyond basic emotions". In this Issue.

Nicholson, A.E. and Brady, J.M.(1994). Dynamic Belief Networks for discrete monitoring. *IEEE Transactions on Systems, Man and Cybernetics, 24, 11,*. Also available at http://citeseer.nj.nec.com/update/106040

Oatley, K. and Johnson-Laird, P.N.(1987). "Towards a Cognitive Theory of Emotions". *Cognition and Emotion, 29-50,*

Ortony, A., Clore, G.L. and Collins, A. (1988). "The cognitive structure of emotions". *Cambridge University Press.*

Ortony, A. (1988), "Subjective importance and computational models of emotions." In V. Hamilton, G.H. Bower & N.H. Frijda, Eds. *Cognitive perspectives on emotion and motivation.* Kluwer, pp. 321-333.

Ortony, A. (2001) On making believable emotional agents believable. In R. Trappl, P. Petta, and S. Payr, editors, Emotions in Humans and Artifacts. The MIT Press.

Prendinger, H., Descamps, S. and Ishizuka, M. (2002). "Scripting affective comunication with life.like characters". In *Applied Artificial Intelligence.* Special Issue on 'Merging cognition and affect in HCI'. F. de Rosis (Ed).

Pearl, A.J (1988) "Probabilistic reasoning in intelligent systems: Networks of plausible inference". *Morgan Kaufman Publishers.*

Picard, R. (1997). "Affective Computing. *The MIT Press, 1997*."

Staller, A and Petta, P (2001). "Introducing emotions into the computational study of social norms: a first evaluation." *JASS*, 4, 1.

Sillince, J.A.A. and Minors, R.H. (1991). "What makes a strong argument? Emotions, highly-placed values and role-playing." *Communication and Cognition*, 24, 3-4, pp.281-298.

---

[i] An ongoing European Research Project that is financed in the scope of the Information Societies Technology Programme IST-1999-29078. The Project includes the following partners: ICCS, University of Edinburgh, UK; DFKI, Germany; SICS, Sweden; AvatarMe, UK; DIS, University of Roma 'La Sapienza', Italy; Dipartimento di Informatica, University of Bari, Italy.

[ii] Compatibility of goals is defined according to psychological plausibility: for instance, it is infrequent to attach a *very high* importance to 'dominating others' and to 'desiring their good', at the same time. Medium weights to both goals might be, on the contrary, plausible.