

From Greta's Mind to her Face: Modelling the Dynamics of Affective States in a Conversational Embodied Agent

Fiorella de Rosis¹, Catherine Pelachaud², Isabella Poggi³,
Valeria Carofiglio¹ and Berardina De Carolis¹

¹ Dipartimento di Informatica, Università di Bari, Italy

² Dipartimento di Informatica, Università di Roma, La Sapienza, Italy

³ Dipartimento di Scienze dell'Educazione, Università Roma Tre, Italy

Abstract

This paper describes the results of a research project aimed at implementing a 'realistic' 3D Embodied Agent that can be animated in real-time and is 'believable and expressive': that is, able to communicate with coherency complex information, through the combination and the tight synchronisation of verbal and nonverbal signals. We describe, in particular, how we 'animate' this Agent (that we called *Greta*) so as to enable her to manifest the affective states that are dynamically activated and de-activated in her mind during the dialog with the user. The system is made up of three tightly interrelated components:

- *a representation of the Agent Mind*: this includes long and short-term affective components (personality and emotions) and simulates how emotions are triggered and decay over time according to the Agent's personality and to the context and how several emotions may overlap. Dynamic belief networks with weighting of goals is the formalism we employ to this purpose;
- *a mark-up language* to denote the communicative meanings that may be associated with dialog moves performed by the Agent;
- *a translation of the Agent's tagged move into a face expression*, that combines appropriately the available channels (gaze direction, eyebrow shape, head direction and movement etc). The final output is a 3-D facial model that respects the MPEG-4 standard and uses MPEG-4 Facial Animation Parameters to produce facial expressions.

Throughout the paper, we illustrate the results obtained, with an example of dialog in the domain of 'Advice about eating disorders'. The paper concludes with an analysis of advantages of our cognitive model of emotion triggering and of the problems found in testing it. Although we did not yet complete a formal evaluation of our system, we briefly describe how we plan to assess the agent's believability in terms of consistency of its communicative behavior.

1. Introduction¹

The efforts done in the last twenty years, in natural language interpretation and generation, resulted in a number of applications of immediate and long-term interest, from generating user-adapted multimedia messages to engaging with the user in a conversation about topics of varying complexity: see, for instance, a recent review in (Zukerman and Litman, 2001) and a collection of papers in (Wilks, 1999). It is often claimed that the users' interest for natural language interaction would be greatly enhanced if computers were given a 'human-like appearance': this would contribute to move the metaphor of human-computer interaction from 'using a tool' (as in graphical interfaces) to 'interacting with a friend', an assistant or an artificial being that is there to help you in solving your task or just for fun. This new metaphor

¹ This work was done in the scope of an ongoing European Research Project (MagiCster: IST 1999 29078), which includes the following *partners*: ICCS, University of Edinburgh, UK (Coordinator); DFKI, Germany; SICS, Sweden; AvatarMe, UK; DIS, University of Roma 'La Sapienza', Italy; Dipartimento di Informatica, University of Bari, Italy and the following *sub-contractors*: Dipartimento di Scienze dell'Educazione, University of Roma Tre, Italy; Department of Psychology, University of Reading, UK; Istituto di Fonetica e Dialettologia, National Research Council, Padua, Italy.

would be consistent with the experimentally-grounded theory of ‘Computers are Social Actors’, according to which human-computer interaction might be considered as a form of interpersonal communication, that follows the main rules of interaction of human societies (Nass et al, 1995, 1996 and 2000; Fogg and Nass, 1997; Moon and Nass, 1996 and 1998).

Natural language generation methods consider how the mental states of the speaker (the sender of the message) and of the interlocutor² (the one who should receive the message), influence content, presentation order and style of the message, so that the communicative goal in the specific situation considered may be achieved. When, rather than generating a monolog, a dialog has to be simulated, the generation process should *dynamically* adapt to the changes occurring in the mental state of the two participants during the dialog itself. These may be due either to ‘exogenous’ factors, that is to events occurring in the external environment (typically, what the interlocutor said in his ‘move’³) or to ‘endogenous’ factors, that is to internal changes due to what the speaker herself is saying (for instance, negative emotions triggered when remembering undesirable past events).

A ‘good’ discourse reflects, in some way, its intentional and attentional structure: rhetorical relations insuring its coherence are manifested through linguistic markers (Mann et al, 1989); a communicative subgoal may be recognised in every text span, the communicative acts reflect social and role relationships between the two interlocutors, and so on. Finally, the ‘affective’ aspects of communication are expressed in more or less overt form (Hovy, 1988; Walker et al, 1997; de Rosis et al, 1999; de Rosis and Grasso, 2000). We include in affective aspects of communication not only ‘basic’ emotions and mood, but also those aspects of the mind that are often considered as ‘extrarational’ in the artificial intelligence community: for instance, ‘doubt’ ‘cognitive overload’ and ‘deception’ (Carberry and Schroeder, 2001; Berthold and Jameson, 1999).

Recent studies proved how interesting and difficult at the same time it is, to manifest and recognise these contents in natural language communication (see, e.g. Paiva, 2000). This problem becomes even harder when the (verbal and nonverbal) communication abilities of an Embodied Animated Agent are considered. In this case, achieving *believability* in the Agent’s behaviour is stated as the primary goal of any such system. Although believability is often considered as ‘awkward to define’ beyond its generic meaning of enabling ‘to accept as real’ (Allbeck and Badler, 2001), it is clear that consistency in the Agent’s behaviour (from moment to moment and from situation to situation) is a key factor in achieving it. An Embodied Agent should respond consistently when the same stimulus occurs at different times and should react in similar ways to similar stimuli with all the components of its communication devices (Ortony, 2002).

A second (and maybe less emphasised) requirement concerns *moderation* in the external expression of thoughts and moods and ‘harmonic variation’ of this expression. The Agent should avoid extreme manifestations of its thought, unless this responds to specific needs, and should display specific aspects of its mental state at the ‘right’ moment, by varying them gradually in agreement with the content of her communication. The way these requirements may be implemented depends on the type of character and on the application in which this character acts. Cartoons are ‘excessive’ by their very nature, they systematically (and on purpose) violate the moderation requirement; their reactions to events are immediate and disappear very quickly, to leave place to reactions to new events. The same behaviour does not apply to the case of a *realistic* Agent: the more an Agent aims at being realistic rather than willing to amaze or astonish the user, the more complex its implementation becomes. A ‘consistent’ behaviour cannot be produced, in a realistic Agent, through a ‘scripting’ implementation process, but should reflect a powerful cognitive model of the Agent and an adequately complex reasoning process.

In this paper, we describe how we designed and prototyped a 3D Embodied Conversational Agent (that we named ‘Greta’) that aims at being ‘natural’ in her appearance and behaviour. We aim at designing Greta as an individual rather

² Our agent may make conversation with either the user or another Agent: we will denote both of them with the generic term of ‘interlocutor’.

³ We call ‘dialog move’ the set of sentences that one of the dialog participants utters in an individual dialog step.

than a generic Agent, thus enhancing, in the user, the impression to communicate with ‘a specific person’. We try to achieve this goal by representing, in a cognitive model of the Agent (what we call her ‘Mind’), her beliefs, desires, intentions, with the relations linking them and the levels of uncertainty attached to these links; this model is employed to simulate how the Agent reacts (both affectively and rationally) to events⁴ occurring during the dialog. Although our Conversational Agent is domain-independent, we will discuss, in this paper, a particular application: an advisory dialog in the domain of ‘eating disorders’. Our Agent is made up of a ‘Body’ and a ‘Mind’. The Agent’s Mind drives generation of dialog moves and tagging of individual components of these moves according to their *meaning*. The generator of Body has access to a repertoire of *signals* that the Agent may employ in her communication process: in our case, face expressions, head moves, gaze direction and so on. The coupling of meaning and signal defines the way that the various message contents will be expressed by the Agent, by mapping communicative meanings into sets of signals. This mapping has been designed so as to avoid ‘overexpressivity’ (and therefore to respect moderation) and to insure the harmonic variation of expression.

The paper is organised as follows: in Section 2, we introduce the example that will guide illustration of all aspects of our Agent. In Section 3, we give a general description of the Agent’s Body: the communicative functions she is able to manifest and the way that believability in her behaviour is achieved. Section 4 describes the Agent’s Mind: after analysing the relationships among goals, emotions and personality, the advantages of dynamic belief networks in the cognitive modelling of the Agent are brought out. Section 5 illustrates a simulation of the Agent’s behaviour in the example dialog and Section 6 is devoted to describing how the Agent’s Body is coupled to her Mind so to enable independence between the two components. Comparison with related work (Section 7) and discussion about advantages and problems of our method and how we plan to evaluate the quality of result produced so far (Section 8) conclude the paper.

2. An example

Let us consider, as we said, an example in the domain of eating disorders. The task Greta has to perform, in this case, is to provide users with information about this kind of problems and eventually suggest how to overcome them. The category of users that might be interested in this type of service are teenagers, possibly with some disorders of affective origin⁵. We believe that these users might feel motivated in discussing their problems with an Embodied Animated Agent, provided this Agent shows a ‘believable’ and ‘engaging’ behaviour: appropriate manifestation of emotions in the dialog is one of the ways believability and engagement may be achieved.

The example dialog that we will consider is one of the simulations we implemented in the scope of the European Project MagiCster: it has a demonstration interest and may be considered as a benchmark of the methods that we will illustrate in the paper. In this dialog, Greta plays the role of the ‘Advisor’ who drives the user towards a positive solution of his or her problem. Greta’s moves are denoted with G0, G1,..., G10, while the user’s moves are denoted with U0, U1,..., U9:

G0	Hi, my name is Greta! I’m here to help you in finding a healthier way of feeding yourself. What’s your name?
U0	<i>My name is Jana.</i>
G1	What’s your problem, Jana?
U1	<i>I've been eating too much for about one year and I realise that I'm not myself, that I'm not thinking quite right!</i>
G2	Oh I’m sorry. That’s a bad situation indeed. Which is your age?

⁴ We define ‘event’ as any change happening in the virtual world, including a dialog move by the Agent’s interlocutor.

⁵ as a matter of fact, several websites already exist in this domain, that are based on traditional hypermedia. For instance: <http://livrite.com/binge.htm> or http://kidshealth.org/teen/food_fitness/binge_eating.html. All websites cited in this paper were last checked on october 2002.

U2 *I'm 17.*

G3 Well: teens are notorious for being hungry frequently, because they need to eat more to support the major growth of their bodies.

U3 *Oh no, it's not a question of increased appetite or overeating from time to time: it's a question of compulsive overeating, in my case. I feel out of control and powerless to stop!*

G4 How bad! Did you already try to do something, to solve this problem?

U4 *Yes, and I did not succeed. I now fear that stopping will not be possible to me! So, I don't know why I'm here!*

G5 Oh no! You have to be confident in yourself! If you want to get up, I may help you.

U5 *What can I do?*

G6 I have some ideas and experience on how to make a smooth, quick transition from overeating back to healthful living. If you wish, I may suggest you what to do.

U6 *Yes, please, I want to make any possible effort to get out of this!*

G7 Fine! This is the way I want to see you! So, let's start, first of all, from cleansing your home from all "binge foods". Whatever your current food downfalls are, get rid of them!

U7 *But I can't stand to throw food away!*

G8 Well, give it to someone, then! And, if your relatives insist on keeping such foods in the house, ask them to keep it away from you for a while.

U8 *That's a good suggestion, I'll try! What else?*

G9 I need to know something more about the way you live, to go on. Do you make any kind of sport, for instance?

U9 *Nothing, since long.*

G10 So, plan to exercise! Choose something you enjoyed doing, even though you may not feel like doing it right now.

.....

After shortly introducing herself, Greta opens the dialog by asking to the user a few informations she needs to orient the dialog. She then reacts to the user moves by showing her that she is 'involved' in her problems and by reassuring her, when needed: to this aim, she shows her emotional engagement at both the linguistic level ("*How bad!, Fine!, etc*) and through appropriate facial expressions. For instance: in the move G2, she shows empathy towards the user; in G3, while providing some background information on teens' eating habits, she emphasises the fact the teens are 'notorious' for being always hungry and need "to eat 'more' to support the 'major' growth of their bodies", by accompanying these adjectives with their corresponding iconic expressions. In move G5, she offers her help to the user and, again, supports this attitude nonverbally. As the user accepts her help, she shows her satisfaction and (at G7) incites her to get rid of binge food; she supports this with an expression of 'proposal', followed by an expression of 'order'. As we saw in these examples, Greta is able to express not only her emotional state but also the kind of performative she employs (suggest, incite etc), some adjectival markers (large) and still more. In Section 3.3, we will give more details on the communicative capacities of Greta. In the following Sections, we will see how emotions are activated in Greta during the dialog, how meanings are associated with every component of the moves generated by the dialog manager and how Greta displays these meanings.

3. The Agent Greta

Greta's believability depends, first of all, on her conversational ability: on one hand, on her ability to engage the user in a dialog that includes the multifaceted shapes of natural dialogs (Wilks, 1999); on the other hand, on the naturalness of her face. Naturalness may have several facets: one of the most superficial ones is the physical aspect. To make Greta's face as natural as possible, we implemented primary and secondary actions (respectively, muscle contractions, wrinkles

and folds). But having a 'nice' 3D model is not enough to get a 'believable' agent: to be believable, the agent should act consistently with her goals, her state of mind and her personality. This means that the agent's behaviour and appearance should be synchronised with her speech and should be made consistent with the meaning of the sentences she pronounces; it should therefore reflect her mental state by avoiding, at the same time, an 'over-expression effect'. Finally, Greta should know when to show and when to hide an expression, according to the situation in which interaction occurs. To this aim, we designed Greta as a BDI (Belief, Desire, Intention) Agent (Rao and Georgeff, 1991) whose mental state includes a representation of the beliefs and goals that drive the feeling of emotions and the decision of whether to display or to hide them. Although, at present, Greta has got only a 3D head, we will refer to her external appearance by talking of her 'Body': this is justified by the fact that we are working to add to her face an upper body that is able to gesture.

3.1. The Agent's body model

The facial model of Greta is quite realistic and appealing. It is the core of an MPEG-4 decoder and is compliant with the "Simple Facial Animation Object Profile" of the standard, which is able to generate the structure of a proprietary 3D model, to animate it and, finally, to render it in real-time (Pelachaud et al, 1996). Two sets of parameters describe and animate the 3D facial model: facial animation parameters (FAPs) and facial definition parameters (FDP). FDPs define the shape of the model while FAPs define the facial actions. When the model has been characterized with FDP, the animation is obtained by specifying for each frame the values of FAPs. The model uses a pseudo-muscular approach, in which muscle contractions are obtained through the deformation of the polygonal mesh around feature points that correspond to skin muscle attachments. Each feature point follows MPEG-4 specifications. A FAP corresponds to a minimal facial action (e.g. move inner_left_eyebrow up, right_lip_corner right). When a FAP is activated (i.e. when its intensity is not null) the feature point on which the FAP acts is moved in the direction signaled by the FAP itself (up, down, left, right). Using the pseudo-muscular approach, the facial model's points within the region of this particular feature point get deformed (see the two examples in **Figures 1a** and **1b**).

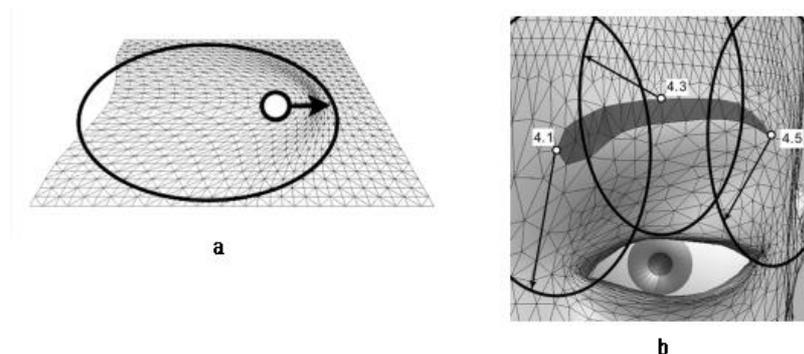


Fig. 1: Muscle deformations in Greta's face

Greta's facial model also includes particular features, such as wrinkles and furrow, to enhance its realism. Brow wrinkles are simulated by discontinuous lines on the forehead and appear when the eyebrows raise (action of the frontalis: **figure 2b**); they have been implemented using bump mapping technique (Blinn, 1978). The nasolabial furrow that is typical of smiling mouths (**figure 2a**) goes from the nose to the corner of the mouth and appears under the action of muscles whose end is around the corner of the mouth. Under the action of the corrugator, vertical bulges between the two eyes are formed, to give rise to a frown (**figure 2c**). Bulges and furrows have been modeled using a specialized displacement function that moves outward points within a specific area. Various kinds of *expressions*, ranging from smile, eyebrow raising, frown and head nod to realistic lip movements, may be generated in this way. At the animation level, the expressions are performed by simulating muscle contractions, as well as extra facial features such as wrinkles and folds.

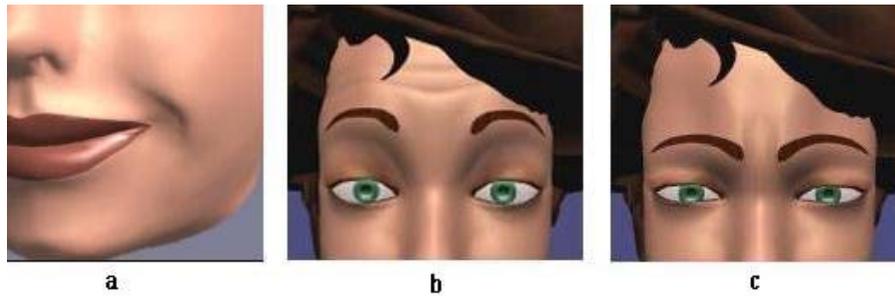


Fig. 2: A smile, an eyebrow raising and a frown in Greta

3.2. Facial signal representation

As explained in the previous Section, to represent the gaze or facial expressions that need to be understood and interpreted by our 3D facial model, we use the Facial Animation Parameters (FAP) as defined by MPEG-4 standard. For instance: a raising eyebrow that marks uncertainty is generated by the FAPs 31 (*raise_left_inner_eyebrow*), 32 (*raise_left_middle_eyebrow*), 33 (*raise_left_outer_eyebrow*), for the left eyebrow and 34 (*raise_right_inner_eyebrow*), 35 (*raise_right_middle_eyebrow*), 36 (*raise_right_outer_eyebrow*) for the right eyebrow. A facial expression is characterized not only by the muscular contraction that gives rise to it, but also by an intensity and a duration. The intensity factor is rendered by specifying an intensity for every FAP. The temporal factor is modeled by three parameters: onset, apex and offset (Ekman, 1979). The onset and offset represent, respectively, the time the expression takes to appear and to disappear; the apex corresponds to the duration for which the facial expression is at its peak intensity value. These parameters are fundamental to convey the proper meaning of the facial expressions, since the same signal may seemingly convey different kinds of information. For instance: raised eyebrows may be part of a surprise expression or may signal uncertainty. Even though these signals involve the same facial expression, they are not identical; surprise is denoted by a very short onset, that is the eyebrow raises very rapidly (Ekman, 1982), while the uncertain expression appears slower on the face. Facial actions may appear more asymmetrically when the expression is faked than when it is felt (Ekman et al, 1981). In our system, every facial expression is characterized by a set of FAPs, an onset and an offset, while the apex (i.e. the duration of the expression) is calculated from the duration of the accompanying speech.

3.3 Meanings the Agent should convey

We integrated the lexicon of face signals with other kinds of meanings that are not generally mentioned in the literature. In fact, it has often been said that facial expressions are a great means to express emotions (Ekman, 1982; Chovil, 1989; Fridlund, 1994; Takeuchi and Nagao, 1993): however, they have a much richer repertoire of communicative functions than bare affective information. They may express doubt; they may have an iconic function, depicting how small or huge is the object or the idea we are mentioning in our utterance; they are linked to the performative of our sentence, and so on. We claim that meanings of facial expressions⁶ may be clustered into four main groups that provide information about:

- the Agent's *beliefs*
- its *intentions*
- its *affective state*
- *metacognitive information* on its mental actions.

The first class includes two kinds of *beliefs*: first-order beliefs about the domain, such as deixis and information on physical or metaphorical properties of referents ('this box is small' or 'this problem is extremely complex'), possibly with an associated degree of certainty, and second-order beliefs about the Agent herself (such as) or about her

⁶ The taxonomy presented here is only related to communicative facial expressions that may occur during speech; we are aware that facial expressions may be linked to other meanings, such as physical state or tics, that we do not include here.

interlocutor. The second cluster gathers facial expressions that are intended to express an *intention* of the Agent: the performative of her sentence (e.g. ‘order’ vs. ‘suggest’), the topic-comment distinction, the rhetorical relation among parts of the message and the turn allocation in conversation. The third cluster represents *emotions*. Finally, the last cluster gathers expressions concerning the kind of *thinking activity* in which the Agent is currently engaged: when ‘remembering a fact’ or ‘trying to make inferences’, we often avoid gaze input by looking up or by closing our eyes, as we are concentrating on our thought and want to avoid information overload. Columns 1, 2,3 in **Table 1** synthesize some of the meanings that may be represented with facial expressions. The reader may refer to (Poggi et al, 2000) for a more detailed description of the communicative functions of Greta.

Communicative functions	Classes	Example Meanings	Signals
Belief	deictic	this, that, here, there	Gaze direction
	adjectival	small, subtle, tiny, difficult	Small eye aperture
		big, long, large, great	Large eye aperture
	certainty	certain	Small frown
		uncertain	Raise eyebrow
Intention	performative	order	Frown, chin up, head straight
		suggest	Raise eyebrow, head aside
	topic/comment	comment	Raise eyebrow
	rhetorical relation	contrast	Raise eyebrow
	turn allocation	Give turn	Gaze direction
Affective state	emotion	joy	Smile, raise cheek
		sadness	Inner raise eyebrow, look down, corner of lip down
Metacognitive	thinking activity	I'm thinking	Gaze direction

Tab.1: Some examples of expressions and their representation with gaze in Greta

3.4. A lexicon for nonverbal expressions

Before visualizing an animation, a ‘dialog move’ has to be translated into a facial expression. To enable this translation, we need to make explicit the correspondence between the meaning of each part of the move and the signal through which Greta’s gaze and face will express it. This correspondence may be defined in terms of a ‘lexicon’, that is a set of (*meaning, signal*) pairs that associate every communicative function with one of the nonverbal modes (gesture, facial expression, gaze, etc) that Greta is able to employ. Let us give some examples of lexical pairs that may be found in the literature, for the four categories of meanings described in the previous Section:

- *Belief*: several studies have shown that raised eyebrow (Cavé et al, 1996; Ekman, 1979) or a sequence of small head nods or gazing at one's interlocutor (Cappella and Pelachaud, 2001) often coincide with the emphasis part of the rheme of a sentence on the nonverbal side, and by a pitch accent at the acoustic level (Hirschberg and Pierrehumbert, 1986).
- *Intention*: by looking at the link between the performatives of sentences and their facial expressions, Poggi and Pelachaud conjectured about the relationship between performatives and affective states. They found, for instance, that the expression of anger (that is characterized by a frown) often accompanies an ‘order’, probably because the speaker will get angry (and has the power to do so!) if what she requested is not done. On the other hand,

‘imploing’ somebody to do something for you is often accompanied by the eyebrow of ‘sadness’, probably because the speaker will feel helpless, and therefore sad, if the Addressee does not perform the requested action (Poggi and Pelachaud, 1998).

- *Affective state*: Paul Ekman and Wallace Friesen described in detail the prototype for the facial expressions of what they consider to be the six ‘basic’ emotions: anger, disgust, fear, happiness, sadness and surprise (Ekman, 1982). Keltner (Keltner, 1996) added to this list the expressions of embarrassment and shame. We add to these expressions those corresponding to the emotions described in the OCC (Ortony, Clore and Collins) model (Ortony et al, 1988).
- *Metacognitive information*: when planning what one is going to say or when trying to remember a fact, one needs to concentrate on one's own thought and, for this reason, often avoids gazing at the interlocutor (Cappella and Pelachaud, 2001; Cassell et al, 1999). Breaking the gaze may be done by looking sideway or up or even by closing the eyes and may have two functions: to signal to the interlocutor that the Agent is thinking or trying to make inferences and to avoid, at the same time, an overload of information coming from the external environment. So, all metacognitive functions are characterized by the act of gazing away from the interlocutor. For some of them, a given direction may be preferred: for instance, looking down when trying to remember some fact and up when trying to make inferences about some fact (Poggi et al, 2000).

Our first step in this direction was to link meanings we wanted to express to combinations of facial signals; that is, we built a *lexicon of facial expressions and gaze*, as a set of rules that map signals into meanings (Poggi and Pelachaud, 1998; Poggi et al, 2000). We based our work on findings from the literature and on the results of analysis of a corpus of videotapes, in which facial expressions and head movements were described using the notation system developed by Poggi and Magno (1996). Some examples of this lexicon may be found in columns 3 and 4 of Table 1.

4. The Agent’s “Mind”

When conversing with a person, our choice of words, facial expressions, gaze, body posture and gesture depends on our interlocutor (are we friends, lovers, enemies; is he my boss or my employee), our view of the interlocutor’s mind (is he able to understand what am I saying, did he go through the same experience I am talking about...), on the situation in which the conversation takes place and finally on our ‘affective state’. However, when talking to a given person, we do not always and necessarily show this affective state: for instance, I will contain my joy if I pass a difficult exam while my best friend did not. When building an Agent, one has therefore to consider on one side the context variables that influence activation of affective states (and, consequently, might trigger a given body expression) and on the other side the factors that may intervene in the (rational or reactive) decision of whether to display this expression.

Greta is set up with a 'background' mental state that drives her initial (and ‘neutral’) expression. This background state is defined according to the factors mentioned above: her relationship with the user, the context in which the conversation takes place and her personality. While her interaction with the interlocutor goes on, the Agent’s mental state evolves and her expression changes, accordingly. Although, as we said, emotions are only a part of the communications we wish to represent, we will focus the description of the cognitive model of our Agent on the way that emotion activation and decay is simulated. Since emotions are partly dependent on personality, we will describe, more specifically, how personality affects the activation of and decay process.

4.1 Goals, emotions and personality

Actions in our life are often part of a plan aiming at some goal: for example, Oetzi, the 5000 B.C. pre-historic man of Similaun, chooses a stone apt to sharpen well and makes a lance for chasing the wildpig successfully. The goals of our everyday plans are not ends in themselves: they all aim in turn at more general goals of biological import that are common to all humans: among them, the biological goals of survival and reproduction and some subgoals like physical well-being, safety, loving and being loved, self-realization, image and self-image. These are terminal goals (TG), ends

in themselves and those to which we assign the highest weight. So much that, if two of them are incompatible (as for instance freedom vs. justice), giving up one of them is a heavy renunciation. With respect to terminal goals, the goals of our everyday life are instrumental (IG), in that they directly or indirectly serve our terminal goals. For instance, chasing the wildpig is instrumental to survival: if the lance is not sharp enough and does not hit the wildpig to death, the wildpig might aggress and kill Oetzi. Instrumental goals are more or less important to us, depending on the strength of their link with terminal goals: at the extent to which an Instrumental Goal is likely to be the only possible means to reach a Terminal Goal, that Instrumental Goal receives a high weight, just because it inherits its weight from the Terminal Goal it serves.

A general consensus exists on the hypothesis that emotions are a biological device aimed at monitoring the state of reaching or threatening our most important goals, be they Terminal or Instrumental (see, for instance, Carbonell, 1980 and Oatley and Johnson-Laird, 1987). Anytime something happens (or is assumed by an Agent to happen) that is likely to produce achieving or threatening a highly weighted goal, the biological device of emotion is triggered, with its whole syndrome of physiological activation, subjective feeling, expressive reaction and behavioral readiness. If, after throwing his lance, Oetzi notices that the stone has not run into the wildpig's heart, the emotion of 'fear' is triggered by his assuming that his goal of survival is challenged: physiological reactions are activated, some of which may appear in the perceivable state of his body (pale face, tremors); and the specific goal of escaping, that might serve the terminal goal of survival, is triggered. Instrumental goals are not linked to the same terminal goals in all contexts. We now feel fear for losing our credit card rather than for a running wildpig, because the goal of survival is linked to other plans than chasing: means-end links between instrumental and terminal goals change with the evolution of culture, technology and, more generally, with the survival strategies adopted in a given environment. A given instrumental goal becomes important (and is likely to cause emotions) if it is the only possible means to reach a terminal goal: if trusting his own lance is not the only way for Oetzi to escape death, because his friends are ready with their lances, he will not feel terror but, say, fear or shame if the lance fails.

This is, in our view, the relationship that holds between goals and emotions: goals, at the same time, *cause* emotions and *are caused by* emotions. They cause emotions since, if an important goal is achieved or threatened, an emotion is triggered: emotions are therefore a feedback device that monitors the reaching or threatening of our high-weight goals. At the same time, emotions activate goals and action plans that are functional to re-establishing or preserving the well being of the individual that was challenged by the events that produced them. So, fear triggers flight, anger triggers aggression; guilt triggers the goal of helping the damaged person or of escaping sanction, and so (Castelfranchi, 2000). Personality is, as well, linked to goals and may be viewed in terms of weights that people put on different goals (Carbonell, 1980; Poggi and Pelachaud, 2000): for instance, sociable persons are ones who give importance to knowing and staying with other people. Of course, the weight attributed to goals may be biologically determined; but, if culture stresses the importance of some goals with respect to others, at least some aspects of personality may be culturally determined.

Since both emotions and personality have to do with the relative importance of goals, is there some link between emotion and personality? Some personality traits may be viewed in terms of the general 'propensity to feel emotions' (Poggi and Pelachaud, 1998, Plutchik, 1980). Picard (1997) calls 'temperament' this subset of personality traits, while other authors relate them directly to one of the factors in the 'Big-Five' model: for instance, neuroticism (Mc Crae and John, 1992). These traits imply, in a sense, a lower threshold in emotion feeling (Ortony, 1988). For instance, a 'shy' person is keener to feel 'shame', especially in front of unknown people. A 'proud' person attributes a high weight to his goal of self-esteem, and in particular to the subgoal of autonomy, of not depending on other people; so, every time one of these goals is achieved, the proud person will feel the emotion of pride. And, conversely, every time these goals are threatened (if, for instance, one is obliged to ask for help), the person will feel the opposite emotion, shame. So: a personality trait (proud) is related to attaching a higher weight to a particular goal (self-esteem, autonomy); and, since that goal is important to that kind of person, the person will feel the corresponding emotion (pride or shame) with a higher intensity.

4.2. Principles in our Modeling of the Agent's Affective State

We now describe how we model the cognitive⁷ reasoning that is involved in triggering and regulating the display of emotions in Greta, while her conversation with the interlocutor goes on. We wish to measure the intensity of emotions and pay particular attention to the way that emotions mix up and change of intensity with time and to how each of them prevails, in a given situation, according to the Agent's personality and to the social context in which the conversation takes place. The following are, more in detail, the *factors we wish to consider in our simulation*:

- *Temperament and personality influence*: the particular factors that, in a given situation, affect Greta's propensity to feel and show emotions and the time they take to decay;
- *Social context influence*: the way that Greta's relationship with the interlocutor and the context in which the conversation takes place influence emotion triggering and display.
- *Dynamics of the Agent's state*: the way that Greta's affective state changes over time as far as the dialog proceeds, as a consequence of events that are due, in our case, to the interlocutor's move;
- *Response decay*: the way that an emotion felt by Greta evolves, in the absence of new specific stimuli, depending on the type of emotion and also on a particular personality trait that affects its 'persistence';
- *Multiple emotions*: the way that several emotions may be activated at the same time and may mix, according to either the 'tub of water' or the 'microwave oven' metaphors suggested by Picard (2000). A user move may trigger one or more emotional reactions, in Greta; once activated, an emotion may last even for several moves. Consequently, several emotions may coexist in a given move, either because they were activated by the same event or because some of them did not yet disappear while the others were triggered.

In addition, we wish Greta to be able to *guess*, if required, the *reason why she feels a particular emotion* and we wish to define the way that expressions competing for the same media combine. Finally, we wish to have, in our model, all the ingredients that enable us to represent, in perspective, *how the Agent's system of goals is revised* when emotions are felt and how this revision influences planning of subsequent dialog moves. We focused our attention, so far, on event-driven emotions (in the OCC model), which correspond to the emotions which are more relevant in our application domain.

4.3. Outline of our model

Greta's Mind is represented with a Dynamic Belief Network (DBN). Belief networks (BNs) are a well-known formalism to represent probabilistic reasoning in directed graphs (Pearl, 2000). DBNs are extensions of BNs, based on the idea that time is divided into *time slices*, each representing the *state* of the modeled world at a particular point or interval in time; this *state* is described by a belief network. Belief networks within time slices are, in general, structurally identical; some of the variables in two consecutive time slices are connected by a link. Variation in the state of the world between two given time slices may affect occurrence of some monitored event. Events, with their causes and consequences, may be described, in their turn, by other subnetworks. 'Observable' nodes enable monitoring both the state of the world and the occurred events (Nicholson and Brady, 1994).

Our model includes several 'static' components that are shown in **Figure 3**:

- $M(T_i)$ represents Greta's mental state at time T_i ;
- $Ev(T_i, T_{i+1})$ represents the event occurred in the time interval (T_i, T_{i+1}) , with its causes and consequences;
- $M(T_{i+1})$ is connected to $M(T_i)$ by links between pairs of nodes that represent, in the two BNs, the belief that a goal will be achieved, in the two time instants. The parameters attached to these links represent the 'degree of

⁷ As in Oatley and Johnson-Laird (1987), "we use the term 'cognitive' to refer to psychological explanations in terms of the representation and transformation of knowledge which may or may not be conscious".

persistence' of this belief. In absence of any event affecting the probability to achieve a particular goal in the considered time interval, these parameters affect the time decay of the emotion related to that goal, until emotional 'equilibrium' is reached;

- $Em\text{-}felt(T_{i+1})$ represents the variation in the intensity of an emotion felt at time T_{i+1} , as a consequence of the change, in Greta's mental state, produced by the events occurred in the time interval (T_i, T_{i+1}) .

We therefore use DBNs as a monitoring system that employs the observation data in the time interval (T_i, T_{i+1}) to generate a probabilistic model of the Agent's Mind at time T_{i+1} , from the model built at time T_i . We employ this model to reason under uncertainty about the consequences of the observed event on the monitored goals.

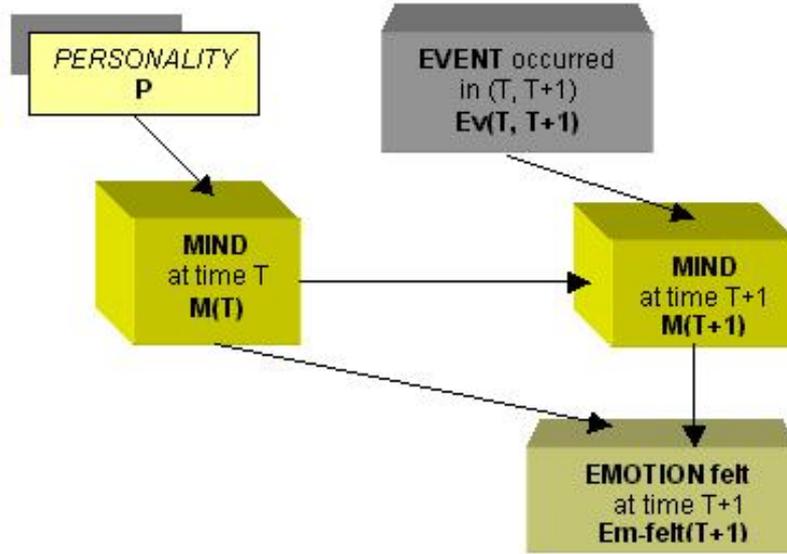


Fig.3: High-level structure of the DBN that simulates emotion triggering in Greta

4.4 Emotion intensity variables and measures

We calculate the intensity of emotions as a function of only two variables:

- *uncertainty* in the Agent's beliefs about the world and, in particular, about the possibility of achieving her goals and
- *utility* assigned to the achievement of goals.

As we said in Section 4.1, emotions are triggered by the belief that a particular goal will be achieved or threatened; the two variables may then be combined, according to the utility theory (Pearl, 2000), to measure *the variation in the intensity of an emotion* as a product of the change in the probability to achieve a given goal, times the utility that achieving this goal takes to the Agent.

Our belief networks represent the relationships among beliefs, in Greta's mental state; they include the following types of nodes:

- *belief nodes*, to represent the Agent's beliefs about the world and about the interlocutor;
- *goal nodes*, to describe the Agent's goals;
- *goal-achievement nodes*, to describe the Agent's beliefs that a particular goal will be achieved or will be threatened: a *weight* is attached to these nodes;
- *event nodes*, to describe the events that might happen as consequences of the interlocutor's move;
- *observable nodes*, to represent observable effects of the events.

If we denote:

- with G Greta, with G_i a goal that Greta wants to achieve, with EV an event that may occur in the considered domain and with e an emotion that is related to achieving or threatening G_i ;

- with Bel G Ach-G_i Greta's belief that G_i will be achieved;
- with P(Bel G Ach-G_i) and P*(Bel G Ach-G_i) the probabilities that Greta attaches to her belief that G_i will be achieved, respectively before and after coming to know that Ev occurred;
- with W(Ach-G_i) the weight that Greta attaches to achieving G_i (in a scale from 0 to 10) and
- with ΔI(e) the variation in intensity of emotion e in Greta, after coming to know that Ev occurred,

then:

- (P*(Bel G Ach-G_i) - P(Bel G Ach-G_i)) is a measure of the change in the probability that G_i will be achieved after Ev occurred;
- ΔI(e) may be calculated as a product of this change of probability, times the weight that achieving G_i takes to the Agent:

$$\Delta I(e) = (P^*(\text{Bel G (Ach-G}_i)) - P(\text{Bel G (Ach-G}_i))) * W(\text{Ach-G}_i).$$

The parameters in the BNs are set so that the probabilities of all goal-achievement nodes are initially set to .5 and that they return to this 'equilibrium value' after the effects of an event have terminated. Therefore, as variations of probabilities range between 0 and .5, ΔI(e) will vary in the range (0,5). By calibrating differently the parameters in the network, domain-specific default values might be established. Default values might be varied, as well, according to the agent's personality and mood: for example, prior probabilities of achieving a goal might be higher for *optimistic* agents. This can easily be obtained by building several 'Mind' networks (one for every personality or context) and by selecting the network to employ in a specific dialog simulation.

According to the OCC theory (Ortony et al, 1988), 'positive' emotions (happy-for, hope, joy etc) are activated by 'desirable' events and consequently by the belief that some goal will be *achieved*, while 'negative' emotions (sorry-for, fear, distress etc) are activated by 'undesirable' events and consequently by the belief that some goal will be *threatened*. Let us consider, for instance, the triggering of 'sorry-for'. This is a negative emotion and the goal that is involved, in this case, is 'preserving others from bad'. During a conversation, Greta's belief about the probability that this goal will be threatened is influenced by her belief that some undesirable event occurred to the interlocutor. The main variables influencing this probability are the desirability of the event and the probability that Greta attaches to the occurrence of this event and to its consequences. An example of belief network that represents the emotion of 'sorry-for' is shown in **Figure 4**. Although, as we said, the network includes two symmetrical components (at times T and T+1), we show in the figure only the component at time T+1, with links with the goal-achievement node at time T. The network says that Greta (G) will be unsatisfied because of an event Ev occurred to the interlocutor U (Bel G (UnsatisfiedFor G U Ev)) if this event is not desirable to U (Bel G Goal U not(Happens Ev U)) and U is a friend (and therefore Goal G not(Happens Ev U)). This combination of beliefs influences the probability that her goal of 'preserving others from bad' will be threatened (Bel G (Thr-PresFBad U)). If Greta comes to know that some undesirable event occurred to U in the interval (T, T+1), this evidence is propagated in the belief network at time T+1 and the probability that the mentioned goal will be achieved is revised; the variation in this probability triggers the emotion of 'sorry-for'. As we said, the intensity of sorry-for also depends on the weight Greta attaches to this goal: to calculate the intensity of this emotion in a specific situation, a weight has to be attached to this goal-achievement node, depending on a personality trait of Greta that we will call 'being other-centered' (or 'compassionate', as proposed in Carbonell, 1980).

Let us assume that Greta is 'very compassionate': the weight she attaches to achieving the goal of 'preserving others from bad' will be, in this case, very high (=8, for instance). Let us now consider what happens after move U1, in our example dialog. This move is interpreted as

Inform (U G Suffers(EatingDisordes U)) ("U informs G that she suffers of eating disorders").

The event 'U suffers of eating disorders', with its consequences, is represented in a BN and is linked to Mind at time T₂. When propagated into Mind(T₂), this event produces a increase of the probability of node (Bel G (Thr-PresFBad U)) from .5 to .91 and ΔI(sorry-for) =.41*10 = 4.1. If, on the contrary, Greta is 'self-centered' (again, to use Carbonell's terminology: weight = 2), the variation of intensity of sorry-for will be much lower (.82).

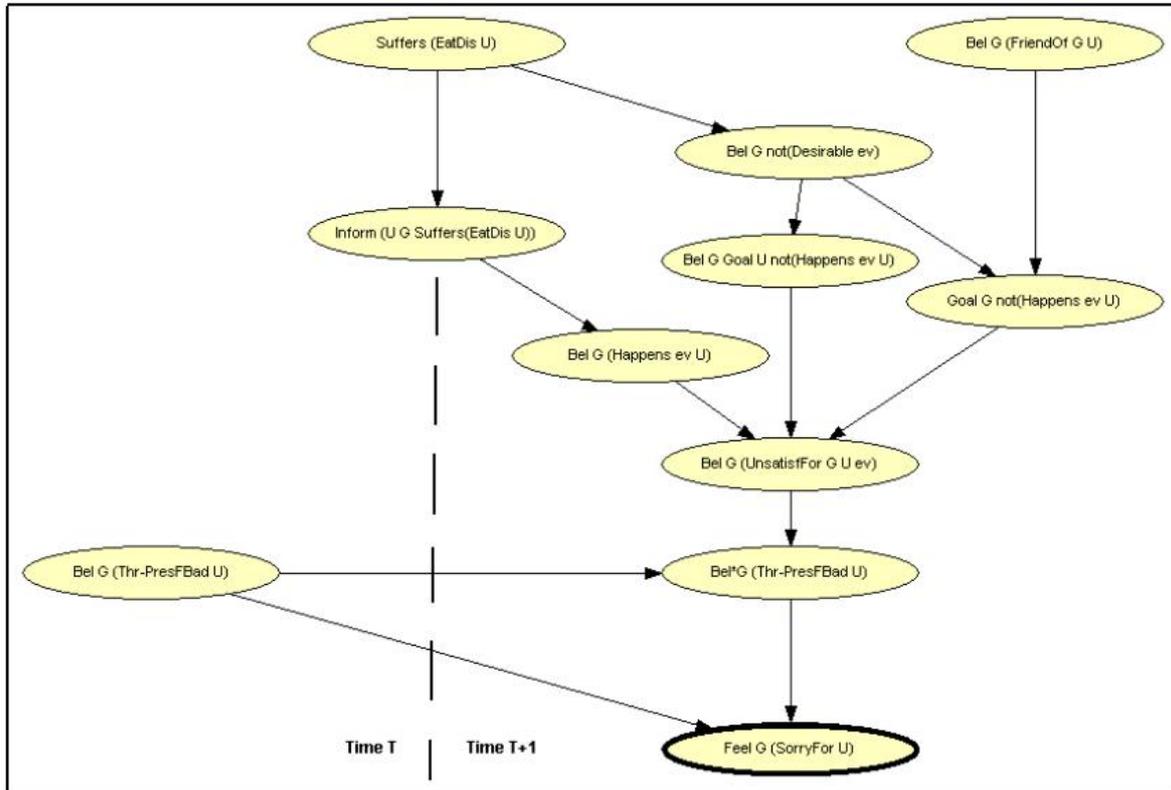


Fig.4: A belief network for triggering the emotion of ‘sorry-for’

4.5 Time decay

The conditional probability table attached to the link between the goal-achievement nodes at two contiguous time instants defines the way the emotion associated with that goal decays, in absence of any event influencing that goal. In our example dialog: as move U2 does not include any ‘emotional effect’, the sorry-for emotion that was activated after move U1 decays exponentially after move U2. By varying appropriately conditional probabilities, we may simulate a more or less fast decay of emotion intensities. For instance: different decays may be simulated for different emotion categories (positive vs. negative, fortune of others vs. wellbeing and so on). We may simulate, as well, different temperaments of our Agent, in which the ‘persistence’ of emotions varies. For instance: a ‘ruminant’ Agent who tends to persist for longtime in the same emotional state or a more ‘impulsive’ Agent, whose permanence in a state is, on the average, shorter.

4.6 Multiple emotions

The Agents’ affective states usually include a mixture of emotions. Rosalind Picard proposes two metaphors for describing the way emotions mix together: in a ‘microwave oven’ metaphor, juxtaposition of two or more states in time occurs even if, at any instant of time, the system is only in one state; in this type of mixing, emotions do not truly co-occur but the affective state switches among them in time: this is, for instance, the case of love-hate. In the ‘tub of water’ metaphor, on the contrary, the kind of mixing allows the states to mingle and form a new state: this is, for instance, the case of feeling wary, a mixture of interest and fear (Picard, 2000). Different emotions may coexist because an event produced several of them at the same time or because a new emotion is triggered while the previous ones did not yet decay completely. Let us assume, for instance, that a ‘desirable’ event occurs to Greta’s interlocutor. Greta will feel, at the same time, happy-for U but also a bit envious, with a relative intensity of the two emotions that depends on her personality and on her relationship with the interlocutor: this is an example of emotion mixing according to the ‘tub of water’ metaphor.

Let us come back again to our dialog example, to see what happens after move U4. As we said, an emotion of sorry-for was triggered after U1 and increased of intensity after U3 because of a new negative event (*U informs G that she suffers of binge eating*). When U says, at move U4, that she might eventually stop the dialog (her sentence “*I don’t know why I’m here*” admits this interpretation), sorry-for is still ‘alive’ in Greta’s emotional state. The new event triggers a new emotion of ‘fear’ which overlaps to the ‘sorry-for’. In this case, the two emotions were triggered at different time instants, while in the example of happy-for and envy they were triggered at the same time. Dynamic belief networks may represent both types of emotion mixing: they may represent how several emotions may be activated at the same time and/or how a new emotion is activated while the previous one did not yet disappear completely (we’ll see some examples in the next Section).

5. A Simulation

We implemented an emotion simulation tool which applies the method described so far, in two different versions. The first version is a testbed with a graphical interface (that we called ‘**Mind-Testbed**’), that we employ to create and test models in various situations. The second one is an executable program (‘**Executable-Mind**’) that may be called from any external program. In Mind-Testbed, belief networks representing the Agent’s initial mental state and the events that might occur in the application domains of interest may be created with the underlying tool Hugin (Hugin, 1988). A set of Greta’s personalities and of contexts in which to test the models may be defined, as well. **Figure 5** shows how the graphical interface of Mind-Testbed may be employed by a System Designer in testing the model in a specific application domain. A personality for Greta and a context are selected, first of all, through icons in the main toolbar. An application domain, an initial mental state for Greta and a threshold for emotion intensities are then set up. After this preliminary phase, the events that correspond to the user moves are introduced, in sequence. The graphical interface shows, in the leftside top frame, the simulation conditions. In the example in the figure: personality=other-centered&optimistic; context=friendship; domain=emotional eating; threshold = 3% (as we wanted to show, in the example, also emotions with very small intensities). In the rightside frame, the Dynamic Belief Network is displayed, with evidence introduced into nodes in different colors. In the leftside bottom frame, the dynamics of emotion intensities is displayed in table form (as in Figure 5) or in graphical form (as in **Figure 6**). In Executable-Mind, when the calling program (in our case, the Dialog Manager) inputs a user move, this is interpreted as a combination of ‘communicative acts’. Executable Mind then analyzes the various acts as events that might activate emotions in Greta. Evidence related to the observed event is introduced into the dynamic belief network, the emotion intensity table is updated with a new column at the considered time instant and these values are sent back to the calling program (again, in our example, the dialog manager).

Let us discuss, in particular, what happens exactly in the example dialog described in Section 2. Initially, Greta is in a ‘neutral’ emotional state (Time 0 in figure 5).

Move U0 does not produce any emotional reaction in Greta (Time 1).

Move U1 is interpreted, as we said, as “*Inform U G (SuffersEatingDisorders U)*”. This is an undesirable event occurring to U that activates an emotion of ‘sorry-for’ in Greta (Time 2).

Move U2 does not produce any new emotional reaction: sorry-for decays (Time 3).

Move U3 is interpreted as “*Inform U G (SuffersBingeEating U)*”. This undesirable event increases again the emotion of ‘sorry-for’ (Time 4).

Move U4 is interpreted as “*Inform U G (EvStopDialog U)*”: U might stop the dialog. This is, again, an undesirable event to G that might occur in the future and therefore provokes a light emotion of ‘fear’ (a kind of ‘concern’) in Greta (Time 5).

Move U5 does not produce any emotional reaction: sorry-for and fear both decrease at Time 6.

Move U6 is interpreted as “*Inform U G (EvStopOvereating U)*”, that is “*the user manifests some intention to interrupt compulsive overeating*”. This is a desirable event to U that activates an emotion of ‘hope’ in Greta (Time 7).

Move U7 does not produce any emotional reaction; all emotions decrease (Time 8).

Move U8 is interpreted as “*Inform U G (LikesSuggestion U)*”, that is “*The user seems to appreciate Greta’s help*”. This is a desirable event to Greta that produces an emotion of ‘joy’ (Time 9)

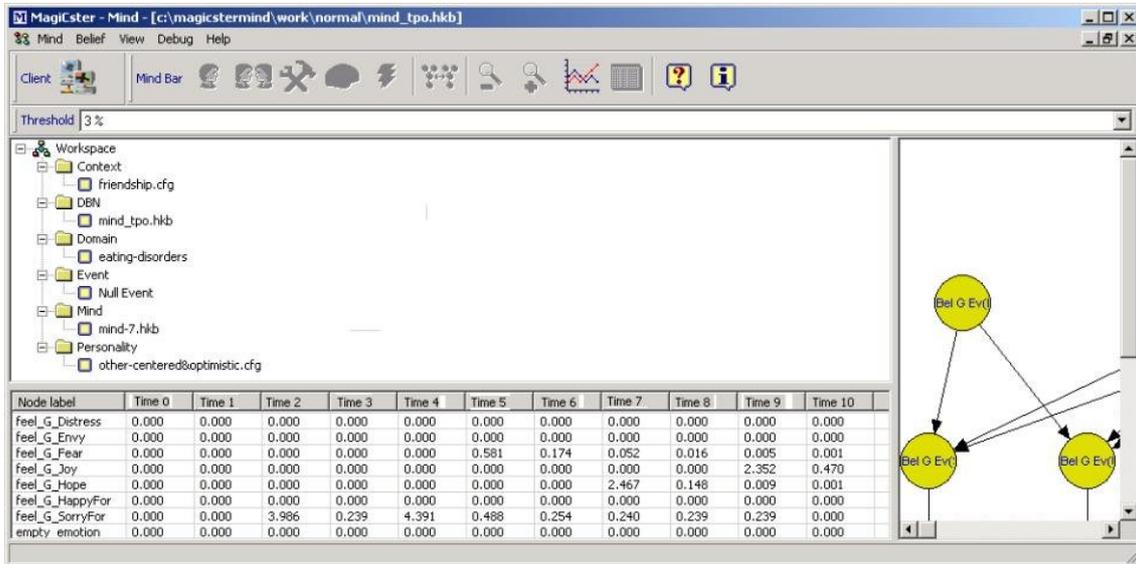


Fig.5: The graphical interface of ‘Mind-Testbed’

As we said, the emotions that have been activated in a given time instant decay progressively, at subsequent times, in a way that depends on the probability table attached to the link between goal-achievement nodes.

The four graphs in **Figure 6**, that have been produced with Mind-Testbed, represent the situation described in our example dialog, with different personalities for Greta. The context is always set to ‘friendship’ and the threshold to 3%, while the following situations are simulated:

- a. Greta ‘*self-centered & pessimistic*’: the Agent associates a high weight to the goals of ‘*preserving self from (immediate or future) bad*’, which activate (respectively) distress and fear. A low weight is attached, on the contrary, to the goals of ‘*achieving the good of others*’, ‘*preserving others from bad*’ and ‘*achieving the future good of self*’, which may activate (after negative or positive events) sorry-for, happy-for or hope.
- b. Greta ‘*other-centered & optimistic*’: the Agent associates a high weight (=10) with the goals of ‘*achieving the good of others*’ and ‘*preserving others from bad*’, which are responsible for activating the emotions of happy-for and sorry-for. She attaches a high weight also to the goals of ‘*achieving the (immediate or future) good of self*’, which activate joy and hope. On the contrary, a low weight (=2) is attached to the goal of ‘*dominating others*’ and ‘*preserving self from bad*’, which (respectively) activate envy and fear.

We introduced a second variable in the simulation conditions, by modifying the parameters that define the time decay of emotions:

- c. in the ‘*ruminant*’ condition, emotions decay slowly;
- d. in the ‘*impulsive*’ condition, they decay faster.

By combining personality traits and time decay factors, we get four different situations:

- A. ‘*self-centered & pessimistic & slow-decay*’

- B. 'self-centered & pessimistic & fast-decay'.
- C. 'other-centered & optimistic & slow-decay'
- D. 'other-centered & optimistic & fast-decay'.

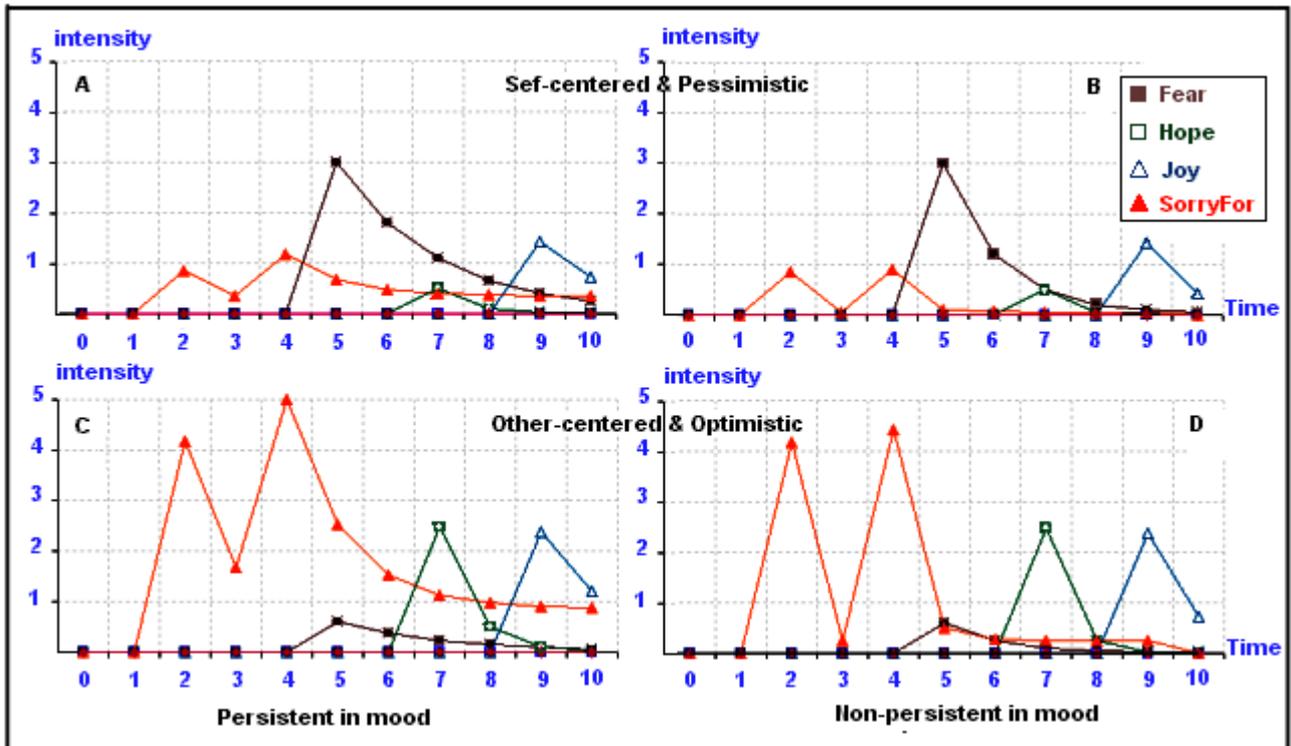


Fig.6: Evolution of the emotional state of Greta in the 10 moves of the example dialog: emotions felt by the Agent at moves G0, G1, ..., G10 are displayed, in the graphs, at times 0, 1,...,10.

This figure shows that sorry-for and hope have a higher intensity in cases C and D, when Greta is 'other-centered & optimistic', while fear is higher in cases A and B, when Greta is 'self-centered & pessimistic'. The intensity of joy is the same in all cases. In cases A and C, time decay of all emotions is slower than in B and D. The result is that:

- in cases A and B, a mild sorry-for is activated at T2, T4, a higher fear at T5 and a mild joy at T9; at T7, the effect of hope is lower than the intensity of other co-occurring emotions ;
- in case C, sorry-for is high at T2 and T4, hope is mild at T7 and overlaps with sorry-for, while fear has a lower intensity than other, co-occurring emotions;
- in D, three emotions are felt neatly during the dialog (sorry-for, hope and joy), while fear has a low intensity, although it is less masked by sorry-for than in C.

The emotion trends shown in the four situations appear to be consistent with the personality trait configurations. By varying appropriately other parameters in the model, we may simulate other situations: for instance, the case in which the dialog occurs in a 'less friendly' context, because Greta notices that the user is joking and is not really interested in getting advice about his or her problem. By increasing the activation threshold for all emotions, we may simulate, as well, a Greta who is rather cold in feeling all emotions or who, though feeling them, does not display them.

6. From dialog moves to Greta's behaviors: linkage between "Mind" and "Body"

Conversations with Greta are implemented with a dialog-simulator in TRINDI (TrindiKit website): the Agent's *dialog moves* are generated, in this tool, as strings in a language that her Body is able to interpret and render appropriately. The input to the Body responds to the following needs:

- it is a *natural language utterance*: it may consist of a unique sentence or a combination of sentences whose coherence is ensured by ‘linguistic markers’ that show the rhetorical relations between them (Mann et al, 1989);
- it includes a description of the *communicative information* associated with those parts of the utterance in which appropriate face expressions should be displayed.

We defined a mark-up language that enables us to associate a semantics to items of interest in the utterances. This gives us the opportunity of keeping our Agent’s Mind independent of her Body, so that the way that dialog moves will be rendered may be selected according to the context in which the system will be employed: for instance, we might use a 2D character in handheld computers or a 3D character with a culture-tailored physical aspect when interaction is with a PC or a large screen. In the next Sections, we will describe in more detail the principles behind Mind-Body coupling.

6.1 The mark-up language

The semantic of the dialog move components is specified through an ‘Affective Presentation Markup Language’ (APML: De Carolis et al, 2002). Tags in this XML-based language correspond to the communicative functions that we listed in Table 1. A dialog move may consist of an individual performative or may correspond to a more complex ‘discourse plan’: for instance, describe the advantages of healthy eating. The input to Greta’s Body is therefore a APML string, in which the natural language discourse is enriched with the tags that the graphical and speech generation components need to produce the required expressions: turn-allocation type, performative, rhetorical relations between communicative acts, deictic or adjectival components, certainty values, metacognitive expressions and finally, emotion intensities. By tagging the text, we therefore provide to Greta an input that enables her to show a consistent verbal and nonverbal expression of every move. The following are the meanings that are associated automatically with moves in our example dialog when Greta is ‘other-centered & pessimistic’ (in addition to several ‘topic-comments’ tags to emphasize particular words, that we omit here for simplicity reasons):

- G0 Hi, my name is Greta! I’m here to help you in finding a healthier way of feeding yourself.
 What’s your name?
 Expressions: a take turn and a performative of ‘greet’, followed by a ‘question’.
- G1 What’s your problem, Jana?
 Expressions: a performative of ‘question’.
- G2 Oh I’m sorry. That’s a bad situation indeed. Which is your age?
 Expressions: a strong emotion of ‘sorry-for’ an adjectival ‘large’ associated with the word ‘bad’ and a ‘question’.
- G3 Well: teens are notorious for being hungry frequently, because they need to eat more to support the major growth of their bodies.
 Expressions: a performative of ‘recognize’ (associated with ‘Well’), an ‘inform’, an adjectival ‘large’ associated with the ‘notorious’ and the ‘more’ word and belief relations of ‘cause-effect’ and ‘justification’.
- G4 How bad! Did you already try to do something, to solve this problem?
 Expressions: an emotion of ‘sorry-for’, a performative of ‘ask’ and an adjectival of ‘difficult’ associated with ‘this’.
- G5 Oh no! You have to be confident in yourself! If you want to get up, I may help you.
 Expressions: a light emotion of ‘fear’, performatives of ‘announce’ and ‘propose’ and an ‘uncertain’ associated with ‘if you want to get up’.
- G6 I have some ideas and experience on how to make a smooth, quick transition from overeating back to healthful living. If you wish, I may suggest you what to do.
 Expressions: performatives of ‘inform’ and ‘suggest’ and several adjectival tags.
- G7 Fine! This is the way I want to see you! So let’s start, first of all, from cleansing your home from all binge foods! Whatever your current food downfalls are, get rid of them!

... and so on.

In **Figure 7**, we show how move G7, in particular, is annotated. The example starts with a ‘take-turn’; then, three types of performatives are employed (confirm, propose and order). An expression of ‘happy-for’ is associated with the sentence in which Greta says to the user that she approves her decision. A tag is associated with the adjective ‘binge’, to enable emphasizing this term through an appropriate facial expression. Topic-comment tags are attached to other parts of the move that have to be emphasized.

```
<turn-allocation type="take turn">
  <affective type="happy-for">
    <performative type="confirm"> Fine!
      <topic-comment type="comment"> This </topic-comment>
        is the way I want to see you!
      </performative>
    </affective>
    <performative type="propose">
      So, let's start, first of all, from cleansing your home from
        <topic-comment type="comment"> all </topic-comment>
        <adjectival type="large"> binge </adjectival> foods!
      </performative>
    <belief-relation type="conditioning">
      Whatever your current food downfalls are,
    </belief-relation>
    <performative type="order">
      <topic-comment type="comment"> get rid </topic-comment> of them!
    </performative>
  </turn-allocation>
```

Fig.7: An example of annotated dialog move

6.2 The Animation

To synthesise the speech part of dialog moves, we use Festival, the speech synthesiser developed at the University of Edinburgh (Black et al, website). The output of Festival provides the input to the face player: to compute the lip movements, we need to decompose the discourse into its list of phonemes with their duration and this is easily done using the Festival’s programming facility. As the text is parsed within the graphics engine, two sets of information are important: the list of phonemes and the list of tags. The lip movement is computed using a three-step algorithm (Pelachaud et al, 1996) that considers the problem of coarticulation, while for each tag we look in our lexicon for the corresponding facial signals. Each of these expressions is represented by a set of FAP values. Several tags may span over the same text segment, meaning that several facial expressions should occur on this segment. In the example in Figure 7, a performative tag with the value ‘order’ and a topic-comment tag with the value ‘comment’ co-occur in the text ‘get rid’. ‘Order’ is shown in the eyebrow region by a frown, while ‘comment’ by a raise eyebrow. These co-occurring communicative functions therefore produce a conflict and the face player has to establish whether the value for the eyebrow should be a frown or a raise eyebrow, because simply adding these two signals would not produce a believable expression. Indeed, adding a frown and a raise eyebrow on our facial model would produce an oblique eyebrow, which is the typical eyebrow characterising the emotion of ‘fear’. So, if we simply add the two signals, we introduce a new meaning that was not intended in the text and is not coherent with it. To solve this kind of conflicts, we have elaborated a method based, again, on Belief Networks (Pelachaud and Poggi, to appear). The nodes of this BN represent communicative functions and facial signals; when evidence about communicative functions to display is

introduced in the network, the signals with the highest priority may be observed. In our example, if ‘order’ and ‘comment’ are introduced at the same time, the output with the highest probability for the eyebrow region is frown: this means that, in this case, the signal of the performative of ‘order’ prevails.

Let us come back, once again, to our example dialog, to describe some of the facial expressions displayed by Greta in the various moves. In move G2, Greta displays the expression of being ‘sorry-for’ the user: inner raise eyebrow (the eyebrow of sadness) and head aside. In her next move, G3, she emphasises the fact the teens are ‘notorious’ for being always hungry with a raised eyebrow and a small head nod. Greta continues her explanation by using adjectival markers “to eat ‘more’ to support the ‘major’ growth of their bodies” and accompanies these adjectives with their corresponding iconic values: that is, by opening her eyes widely as to denote something ‘big and large’. In move G5, Greta supports her offer of help to the user with raised eyebrows and head aside, to show that she is uncertain on whether the user will accept this offer. As, in the next move, the user does accept it, she incites her to get rid of binge food by displaying a smile and bending her head forward towards the user.

Figure 8 shows some examples of these expressions: a ‘neutral’ expression, a ‘sorry-for’, an ‘order’ and a ‘large’ expression. These are only ‘static’ samples of the dynamic expressions that appear on our agent’s face while the dialog goes on.

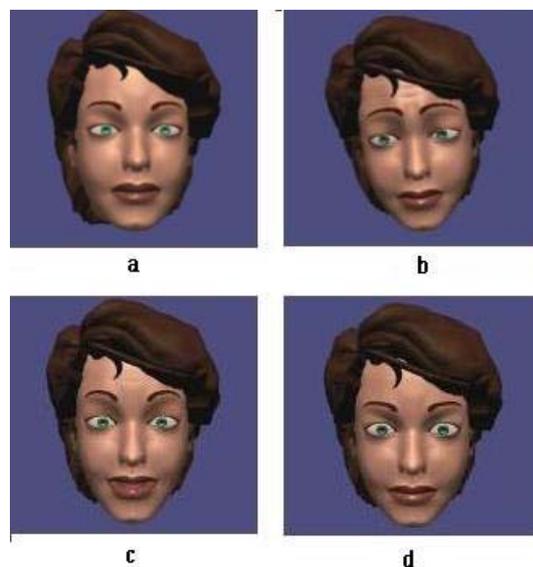


Fig.8: Some expressions of Greta’s affective state: ‘neutral’(8a), ‘sorry-for’(8b), order (8c) and ‘large’ (8d)

7. Related work

Emotions and personality are considered, since long, essential ingredients of believability in Embodied Animated Agents (Bates, 1994; Loyall and Bates, 1997; Arafa et al, 1999; Elliott et al, 1999; Isbister and Nass, 1999; Taylor et al, 1999; André et al, 2000; Lester et al, 2000; Marsella et al, 2001). Our cognitive model of emotions builds on these experiences and, more in general, on the contributions of those who wondered whether and how computers may be endowed with the ability to simulate feeling and display of emotions (Frjida and Swagerman, 1987; Elliott, 1994; Hudlicka and Fellows, 1996; Pfeifer and Nicholas, 1985; Pfeifer, 1988, in addition to those we have already cited in other parts of this paper). The main difference between Greta’s mind and these systems is in the ‘grain size’ of knowledge represented and in the theory applied to measure the emotion intensities. Rather than representing directly the effect of events, objects and actions on the Agent’s affective state by an application of OCC theory, 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. This enables us to achieve the purposes mentioned in Section 4.2:

mainly, representing the dynamics of affective states, the mixing-up of emotions and the prevalence of different emotion mixtures, in the same situation, in Agents with different personalities and in different contexts.

The ideas we apply about the relationships among goals, personalities and emotions are not entirely new. Bates and Reilly said that “Emotions have intensities that are influenced by the importance of the goal that generated them” and that each emotion has its intensity lowered with time, until the intensity is zero or another stimulus happens (citation from Picard, 2000). 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, 1987). Roseman’s theory establishes a relationship between ‘appraisals’, that are defined in terms of motivational and situational state, and emotion raising; ‘problem type expectedness’ and ‘control potential’ are represented, in this theory, in terms of beliefs, and therefore are in close relationship with the BDI model that we employ in Greta’s mind (Roseman, 1991). Picard suggests that affective computational models should include a representation of ‘response decay’ (“an emotional response is of relatively short duration and will fall below a level of perceptibility unless it is reactivated”) and of the influence of personality and temperament. In Prendinger and Ishizuka’s appraisal process, events are evaluated as to their emotional significance for the Agent. Emotions are activated by elicitation rules, without uncertainty; the Agent’s preferences and goals are introduced in the eliciting conditions in the form of ‘likes’ and ‘wants’ predicates, each with its intensity. The emotion intensities are calculated as a logarithmic combination of the sum of exponentials of the measures attached to some of the variables in the rule conditions. For instance: the intensity of *joy* coincides with 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. Although the meaning of this function is different, its form is identical to the form of the function that is employed in Reilly and Bates’ Em, to compute the Agent’s *mood* as a combination of its positive and negative emotions (Picard, 2000). If the number of variables in the domain of this function is low, the function selects, in a way, the highest value assumed by these variables: therefore, the ‘strongest’ eliciting condition gives a measure of the intensity of the emotion triggered. Personality does not influence emotion elicitation, but only the ‘filtering’ process, in which it is decided whether the emotion should be expressed or suppressed. In this case, the ‘degree of friendliness’ with the interlocutor combines additively with ‘social power’ and ‘social distance’ between the Agent and the interlocutor and with the emotion intensity, to decide whether the emotion should be suppressed or not (Prendinger and Ishizuka, 2001).

In Affective Reasoner, Elliott examines the variables that may influence the intensity of an agent’s affective state, by proposing a scale and a default value for each of them (Elliott and Siegle, 1993). That paper was a remarkable step forward of the theory previously proposed by Ortony (1988): although the need to define an ‘emotion-intensity calculation function’ (either general or emotion-specific) was mentioned by the two authors, no specific proposal for the form of this function was made. The core of Elliott’s system of variables is the need to define a precise meaning and measuring method for each of them; this is an essential prerequisite for defining a function that combines the various variables in a measure of intensity of an emotion. However, combining several heterogeneous measures in a unique function is not easy, for several reasons. First of all, when different scales are employed for each variable, 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 it should measure (in our case, the emotion intensity), with the consequent need of ‘normalising’ the obtained value or of ‘cutting’ values out of this range. The computational model of emotion that was proposed in Affective Reasoner has been applied recently to model emotional reasoning in Emile (Gratch, 1999 and 2000). 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 probability of goal attainment. The new and interesting point is that this last variable is measured from representation of the plans that could bring that goal about. Emile was incorporated in the Mission Rehearsal Exercise system (Gratch and Marsella 2001), in which it was integrated with Steve (Rickel and Johnson, 1999). This system contains also some components of the Interactive Pedagogical Drama: those that define the impact

of an emotion on the nonverbal behaviour of the agent and those that define transitional states between different emotions.

What is new, in our modelling method, is our application of utility theory and our representation of the appraisal process in terms of dynamic belief networks. We only employ two measures (uncertainty of beliefs and utility of goals), while Elliott proposes a different measure for each variable. Elliott's 'importance of achieving a goal/not having a goal blocked' directly corresponds to the 'weights' we attach to goal achievement or threatening. All variables associated with our 'belief' nodes are a representation of Elliott's '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 proposes), we discretise 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. Our emotion intensity calculation function is therefore a pure application of utility theory to the simulation of emotion elicitation.

Belief networks have already been applied to represent external expression of emotions (Ball and Breese, 2000). Dynamic Belief Networks provide some additional advantages in modelling emotion triggering and display, as they enable representing the following crucial aspects of this problem:

- personalities are linked to weights given to goals, as originally proposed by Carbonell (1980) and, more recently, by Ortony (2002): this enables defining a trait for every goal and combining traits into consistent personalities. Other traits are simulated by varying the threshold for emotion triggering and the parameters defining the time decay of every emotion;
- emotions are triggered by variations in the belief that a given goal can/cannot be achieved (a typically dynamic mechanism);
- these variations of belief may be originated by 'external events' (for instance, the interlocutor's dialog moves) or by 'internal-endogenous events' (for instance, the content of own move);
- emotions may activate new goals, in either a rational or a reactive way;
- all the mentioned relationships are governed by uncertainty.

In addition, the same formalism may be applied to simulate the emotion *regulation* mechanism, according to which the Agent may 'feel' an emotion but may decide not to show it, and the *expression* mechanism, in which possible conflicts among expressions originating from different meanings associated with a given message are resolved.

A final word about the independence of the Agent's mind from her body. As we anticipated in Section 6, separation of the two components enables rendering the output of the Agent's mind (dialog moves with associated meanings) in a way that depends on the physical and cultural context. This is an exciting perspective, if we wish our Animated Agents to be able to travel along different virtual (and natural) worlds, by dynamically adequating their aspect to the interaction device available and to the cultural and social norms of the world in which they live or land.

8. Towards Evaluation and Further Improvements of Greta

We will now discuss advantages and problems of our affect simulation method and will add some considerations about scalability and evaluation perspectives and criteria.

8.1. Advantages

One might wonder why we didn't apply directly the two theories that are prevailing in the design and implementation of today's believable animated agents: the Big Five model of personality and the OCC theory of emotions. The reason why we did not adopt the Big Five model is that we wanted a finer differentiation in the Agent's behaviour tendencies than the extrovert/introvert, friendly/unfriendly and so on, that are proposed in that model: we wanted to represent at least in part the underlying variables of personality traits, in particular the reasons why Agents with different traits typically feel emotions differently (Moffat, 1997). Defining personalities in terms of behavioural dispositions that originate from the Agent's belief that her motivations will be achieved or threatened provides us such an opportunity. Representing in a cognitive model the beliefs and goals that brought our Agent to feel a given emotion enables us, more in general, to have some insight into the underlying mechanism of her -rational and emotional- behaviour. Someone might argue that we employ this knowledge just for shallow behavioural manifestations (to display emotions, with varying degrees of intensities, on the Agent's face) and that this does not consider the impact of personality on deeper components of her behaviour, like planning and action performance. This is only apparently true. This paper was focused on Greta's display of emotions and on linkage between her mind and her body: this is the reason why we limited our description to the more shallow (expressive) aspects of her affective behaviour. At the same time, however, we are working on emotional dialog simulation, to study how planning of dialog moves is influenced by the Agent's beliefs and goals and by their long and short-term (personality and emotion-based) variations. This way, the interest of a BDI representation of Greta's mind becomes more clear, also considering that using a unique mental model to generate different aspects of her behaviour will make us more confident about consistency of this behaviour. In particular, given a communicative goal, we may be sure that *what* Greta says in the dialog is in agreement with *the way that* she says it.

Finally, we have to add, to the mentioned advantages of a cognitive model of emotion triggering, a second (and non-secondary) advantage. This is in the possibility of going towards a *componential* approach to emotion expression in which, rather than displaying stereotyped emotions in Greta's face, some distinguished components of her mental state (beliefs, goals and so on) are directly displayed into specific facial features (Smith and Scott, 1997). We anticipated some first results in this direction in Section 5.2, when we described how we deal with co-occurring (and potentially conflicting) expressions.

8.2 Problems

As we said, we modelled, so far, a subset of the emotions that are described in the OCC model: the event-driven categories of 'fortune of others', 'well being' and 'prospect-based' emotions (Ortony et al, 1988). The belief network that represents the Agent's mental state in the generic time instant includes about 30 nodes: this number increases considerably with the number of emotions modelled and when two contiguous time slices are combined with an event-network in a DBN. The complexity of our DBNs holds, however, within reduced limits and the time required to propagate evidence and update the emotion intensities is very short, thanks to the effectiveness of the algorithm employed by Hugin (Hugin, 1988). We like emphasising that our Embodied Conversational Agent is completely domain-independent: every aspect of its behaviour is generated, from an initial 'communicative goal' in a given domain. We tested our system in a few 'toy dialogs': in particular, a dialog that reproduces Picard's 'Marathon example' which was appropriate to test emotion mixing (Carofiglio et al, 2002). We then developed two working examples in the medical domain: explanations about drug prescriptions (Pelachaud et al, 2002) and advice about eating disorders, that is described in this paper. These examples were both of interest to test the overall believability of Greta, in domains in which empathy is crucial to make conversation acceptable.

To extend our models to new categories of emotions in the OCC model, we will have to revise the structure of Greta's initial mental state so as to include new 'monitoring' nodes, new beliefs and new goals. In particular, new goal-achievement or threatening nodes for the 'confirmation' category, new 'standard and norm' monitoring nodes for 'attribution' and 'attraction'. Introducing a new domain requires, on the contrary, anticipating the events that might happen in this domain and representing them as belief networks.

We found Mind-Testbed a very effective tool to test the role of context and personality in the simulation of dynamic emotion feeling in dialogs. The following are the main problems that our informal evaluation let us appear:

- a. *Thresholding spurious variations*: building our belief networks required assigning a consistent set of conditional probability distributions to these nodes. When experimental data are not available, subjective estimates have to be done (and we would not see how to measure these parameters empirically, in cognitive modelling of emotions). Network building is therefore an iterative process, in which parameters are tuned after checking the model behaviour in different conditions. In addition, we had to define how goal weights were associated with personality traits. Calibration of parameters is a well known general problem of belief network building: in our case, calibrating the prior and conditional probability tables so as to avoid small, ‘spurious’ variations in the probability of monitored goals was rather difficult. This is one of the reasons why we introduced a threshold, below which an increase in the probability of monitored goal-achievement nodes does not affect triggering of emotions: this ‘threshold effect’ appears to be psycho-cognitively justified, was already proposed in Ortony’s and in Pfeifer’s emotion modelling methods (Ortony, 1988; Pfeifer and Nicholas, 1985) and allows a natural adaptation of emotion triggering to the Agent’s ‘temperament’.
- b. *Over-emotion effect*: a common risk in emotion modelling methods is that, with some personality trait combinations and in some situations, too many emotions are activated at the same time during the dialog: this risks to violate the ‘moderation’ in the external expression of thoughts and moods, that we mentioned (in the Introduction) as a requirement that any ‘realistic’ Embodied Agent should show. In our example dialog, a host of emotions overlap, in a small number of turns. Of course, this depends on how affective the domain and the interaction contexts are: so, in some cases (like the one we consider in this paper), an emotional attitude may contribute to achieving the conversational goal; in others, it might be an obstacle in achieving it and should be avoided. As we showed in figure 6, variation of parameters in our models enable us to get this adaptation in a rather easy and effective way.
- c. *Time stability*: another problem issue is due to the sequencing of different emotions in a dialog: fast changing of emotions of opposite sign from a move to the following one (from sorry-for to fear, from fear to hope, and so on) might produce an overall ‘unbelievable’ behaviour in Greta, an impression of ‘inconsistency’. Again, in some cases emotions have to change, when a fast sequencing of opposite events occurs; this may easily happen in ‘highly emotional’ domains or dialog parts. In other cases, emotion filtering is needed and, again, may be obtained, in our model, by varying the model parameters.
- d. *Emotion masking*: Some situation might require the ‘masking’ of a felt emotion. In (De Carolis et al, 2001) we describe how criteria for establishing whether to display an emotion or to hide it may be formalized with context-dependent rules. We separate emotion triggering from their regulation, so as to simulate the role played by the two interlocutors’ personalities, their relationship, the social interaction context and so on, in regulating emotion expression. Masking an emotion with another one or with a ‘neutral’ expression does not necessarily imply neutralization of the first emotion over the second one: it may imply a variation of its intensity or its duration (decrease, increase). The intensity of an emotion may be decreased by eliminating its expression in some part of the face (for example, in the eyebrow region), by shortening the duration of the expression or by reducing contraction of the muscles involved in the expression. For example: a person may mask his sadness with a smile; however, the face will not display happiness, as sadness might still be shown on the upper part of the face (with the typical oblique eyebrow). A person may also simulate an emotion when he does not feel anything: putting a smile on one's face is the most frequent figure in both simulating or masking cases. Such a smile may appear or disappear very fast and usually tends to be asymmetric (Ekman, 1982).
- e. *Expression overlapping*: which expression should Greta show, when she is supposed to ‘feel’ several emotions at the same time or has to display an emotion and a belief, an intention or a meta-cognitive expression? Literature is full of examples of complex expressions: see, for instance, the variety of ‘smiles’ that may be

observed in humans, in (Ekman and Friesen, 1982). As we said in the previous Section, we developed a method for combining concurrent expressions. In general, signals cannot follow a simple additive rule: for instance, adding the raise eyebrow of ‘emphasis’ to the frown of ‘anger’ produces the oblique eyebrow shape of ‘fear’; adding the two signals produces, in this case, a new signal with a completely different meaning. Additive rules make no sense, in particular, for the signal ‘gaze direction’; when several concurrent tags involving gaze direction appear on the same text segment, they produce a conflict. As we said, we built a BN that links the communicative functions to the signals: the choice of the final signal to display is thus obtained through probabilistic reasoning. We are working, now, at a more complex gaze model, in which the speaker’s gaze depends not only on the communicative functions that are specified in the APML tags, but also on what the interlocutor is doing (gazing at the speaker, smiling...) and what the speaker was previously doing (gazing away, smiling etc: Cappella and Pelachaud, 2001).

8.3 Evaluation

The more general problem of our Agent’s *believability* merits a particular comment. In their discussion of how Embodied Conversational Agents might (or should!) be evaluated, Nass and colleagues mention consistency between verbal and nonverbal cues as a potentially important factor of user satisfaction and demonstrate its impact on several measures of preference (Nass et al, 2001). Consistency is, to several authors, a key factor of believability (see, for instance, Ortony, 2002) and inconsistency is an impending danger for Embodied Conversational Agents, that may arise from several mistakes in their design: the content of a move may be inconsistent with its expression, verbal and nonverbal components of communication may be inconsistent, synchronisation of these components may be imperfect, personality traits may be combined in a psychologically implausible way, and so on. No one knows which of these factors merits priority in the design of an ECA. In our case, consistency between verbal and nonverbal components of a move is achieved through automatic tagging of the moves with the APML language described in Section 6.1. This insures, as well, that meanings are associated with move parts consistently all along the dialog. Interpretation of these meanings by Greta’s body in terms of the (meaning, signal) lexicon described in Section 3.4 insures that meanings are displayed consistently all along the dialog. As far as emotion activation is concerned, in particular, our main concern is to be sure that weights are assigned to goals so as to produce consistent personalities. Ortony (2002) suggests applying, to this purpose, theories about clustering of personality traits such as the Big Five model (Mc Crae and John, 1992): although this is a substantial contribution to the definition of an Agent’s personality as a combination of consistent traits, it still leaves open, to us, the problem of association between traits and goal weights.

To assess how ‘believable’ the Agent we developed is, we are designing a set of in-depth evaluation studies, in cooperation with the Department of Psychology at the University of Reading. In these studies, we will focus our attention on the impact of Greta’s consistency into her believability. The test domain will be that of persuasion about healthy eating, in which we will evaluate, in particular, the effect of negative and positive emotion expressions on a set of variables measuring the efficacy of persuasion attempts and will check whether and how introduction of various kinds of inconsistency affects these measures.

Acknowledgements

The anonymous reviewers and the Guest Editor were really keen in orienting us towards a revision of our work (not only of its description), by raising questions to which it was sometimes hard to find a convincing answer. We therefore owe to them a warm and not ritual acknowledgement.

References

- ALLBECK, J.M. & BADLER, N.I. (2001). Consistent communication with control. In C. PELACHAUD & I. POGGI, Eds. *Proceedings of the Workshop on "Representing, annotating and evaluating non-verbal and verbal communicative acts to achieve contextual embodied agents"*. Autonomous Agents '01, Montreal. pp. 21-26.
- ANDRE', E., RIST, T., VAN MULKEN, S., KLESEN, M. & BALDES, S. (2000). The automated design of believable dialogs for animated presentation teams. In J. CASSELL, J. SULLIVAN, S. PREVOST & E. CHURCHILL, Eds. *Embodied Conversational Agents*, MIT Press, Cambridge, Ma. pp. 220-255.
- ARAFA, J., CHARLTON, P., MAMDANI, A., & FEHIN, P. (1999). Designing and building Personal Service Assistants with personality. In S. PREVOST & E. CHURCHILL, Eds. *Proceedings of the Workshop on Embodied Conversational Characters*, Tahoe City. pp. 93-101.
- BALL, G. & BREESE, J. (2000). Emotion and personality in a conversational agent. In J. CASSEL, J. SULLIVAN, S. PREVOST & E. CHURCHILL, Eds. *Embodied Conversational Agents*, MIT Press, Cambridge, MA. pp. 189-219.
- BATES, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37, 7, pp.122-125.
- BERTHOLD, A. & JAMESON, A. (1999). Interpreting symptoms of cognitive load in speech input. In J. Kay, Ed. *Proceedings of User Modeling '99*, Banff, Canada. SpringerWienNewYork. Pp.235-244.
- BLACK, A.W., TAYLOR, P., CALEY, R. & CLARK, R. Festival. <http://www.cstr.ed.ac.uk/projects/festival/>.
- BLINN, J. (1978). Simulation of wrinkled surfaces. *Computer Graphics*, Vol. 12 (3). pp. 286-292.
- CAPPELLA, J. & PELACHAUD, C. (2001). Rules for Responsive Robots: Using Human Interactions to Build Virtual Interactions. In A.L.VANGELISTI, H. T. REIS & M. A. FITZPATRIC, Eds. *Stability and Change in Relationships*. Cambridge University Press, New York. . pp. 325-353.
- CARBERRY, S. & SCHROEDER, L. (2001). Recognizing and conveying attitude and its underlying motivation. In F DE ROSIS, Ed, *Proceedings of the Second Workshop on 'Attitudes, Personality and Emotions in User-Adapted Interaction'*. User Modeling '01, Sonthofen, Germany. pp. 30-35.
- CARBONELL, J.G. (1980). Towards a process model of human personality traits. *Artificial Intelligence*, Vol. 15, pp. 49-74.
- CAROFIGLIO, V., DE ROSIS, F. & GRASSANO, G. (2002). Mixed emotion modeling. In L. CANAMERO and R. AYLETT, Eds, *Proceedings of the AISB '02 Symposium on "Animating Expressive Characters for Social Interaction"*. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, Pub. London. pp. 5-10.
- CASSELL, J., TORRES, O. & PREVOST, S. (1999). Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation. In Y. WILKS, Ed. *Machine Conversations*. Kluwer, The Hague. pp. 143-154.
- CASTELFRANCHI, C. (2000). Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions. In A. PAIVA, Ed. *Affective Interactions*. Springer LNAI 1814, Berlin. pp. 76-106.
- CAVE', C., GUAITELLA, I., BERTRAND, R., SANTI, S., HARLAY, F. & ESPESSER, R. (1996). About the Relationship Between Eyebrow Movements and F0 Variations. In *Proceedings of ICSLP'96: The Fourth International Conference on Spoken Language Processing*. Vol 4, pp 2175-2178. Philadelphia, PA.
- CHOVIL, N. (1989). Communicative Functions of Facial Displays in Conversation. *PhD Thesis*. University of Victoria.
- DE CAROLIS, B., CAROFIGLIO, V. & PELACHAUD C. (2002). From Discourse Plans to Believable Behaviour Generation. *Proceedings of the International National Language Generation Conference*. New York.
- DE ROSIS, F., GRASSO, F. & BERRY, D. (1999). Refining Instructional Text Generation After Evaluation. *Artificial Intelligence in Medicine*. Vol. 17, pp. 1-36.

- DE ROSIS, F. & GRASSO, F. (2000) Affective Natural Language Generation. In A. PAIVA Ed. *Affective Interactions*. Springer LNAI1814. pp. 204-218.
- EKMAN, P. (1979). Human Ethology: Claims and Limits of a New Discipline: Contributions to the Colloquium. In . M. VON CRANACH, K. FOPPA, W. LEPENIES & D. PLOOG, Eds. *About brows: Emotional and conversational signals*. Cambridge University Press. New-York. pp. 169-248.
- EKMAN, P., HAGER, J.C. & FRIESEN, W. (1981). The Symmetry of Emotional and Deliberate Facial Actions. *Psychophysiology*. Vol. 18(2), pp. 101-106.
- EKMAN, P. (1982). *Emotion in the human face*. Cambridge University Press.
- EKMAN, P. & FRIESEN, W. (1982). Felt, false, miserable smiles. *Journal of Nonverbal Behavior*. Vol. 6(4), pp. 238-251.
- ELLIOTT, C. & SIEGLE, G. (1993). Variables Influencing the Intensity of Simulated Affective States. In *Proceedings of the AAAI Spring Symposium on Mental States '93*. pp. 58-67.
- ELLIOTT, C. (1994). Research problems in the use of a shallow Artificial Intelligence model of personality and emotion. In *Proceedings of 12th AAAI Conference*. pp. 9-15.
- ELLIOTT, C., RICKEL, J. & LESTER, J. (1999). Lifelike pedagogical agents and affective computing: an exploratory synthesis. In M. WOOLRIDGE & M. VELOSO, Eds. *Artificial Intelligence Today*. Springer-Verlag LNCS 1600. Berlin. pp. 195-212.
- FESTIVAL website: <http://www.cstr.ed.ac.uk/projects/festival>
- FOGG, J. & NASS, C. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies*. Vol. 46. pp. 551-561.
- FRIDLUND, A. (1994). *Human facial expression: An evolutionary view*. Academic Press. New York.
- FRJIDA, N.H. & SWAGERMAN, J. (1987). Can computers feel? Theory and design of an emotional system. *Cognition and Emotion*. pp. 235-257
- GRATCH, J. (1999). Why you should buy an emotional planner. In *Proceedings of the Agents'99 Workshop on Emotion-based Agent Architectures*.
- GRATCH, J. (2000) Emile: Marshalling Passions in Training and Education. In *Proceedings of the 4th International Conference on Autonomous Agents*. Barcelona, Spain. pp. 325-332.
- GRATCH, J. & 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. pp. 278-285.
- HIRSCHBERG, J. & PIERREHUMBERT, J. (1986). The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*. pp. 136-144.
- HML webpage: New York. <http://www.humanmarkup.org>
- HOVY, E.H. (1988). Affect in text. In *Generating natural language under pragmatic constraints*. LAWRENCE & ERLBAUM Associates, Pubs. pp. 341-381.
- HUDLICKA, E. & FELLOUS, J.M.(1996). Review of computational models of emotions. *Technical Report 9612*, Arlington, Ppsychometrics Associates, Inc.
- HUGIN website: <http://www.hugin.com>

- KELTNER, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion*, Vol. 10(2), pp. 155-171.
- ISBISTER, C. & NASS, C. (1999). Personality in conversational characters: building better digital interaction partners using knowledge about human personality preferences and perceptions. In S. PREVOST & E. CHURCHILL, Eds. *Proceedings of the Workshop on Embodied Conversational Characters*. Tahoe City. pp. 103-111.
- LESTER, J.C., TOWNS, S.G., CALLAWAY, C.B., VOERMAN, J.L. & FITZGERALD, P.J. (2000). Deictic and emotive communication in animated pedagogical agents. In J. CASSELL, J. SULLIVAN, S. PREVOST & E. CHURCHILL, Eds. *Embodied Conversational Agents*, MIT Press. pp. 123-154.
- LOYALL, A B., & BATES, J. (1997). Personality-rich believable agents that use language. *Proceedings of Autonomous Agents 1997*, pp. 106-113
- MANN, W.C., MATTHIESSEN, C.M.I.M. & THOMPSON, S.A. (1989). Rhetorical Structure Theory and Text Analysis. *Information Science Institute Research Report 89-242*. Marina Del Rey, California.
- MARSELLA, S. GRATCH, J. & RICKEL, J. (2001). The effects of affect: Modeling the impact of emotional state on the behavior of interactive virtual humans. In C. PELACHAUD & I. POGGI, Eds. *Proceedings of the Workshop on "Multimodal Communication and Context in Embodied Agents"*. Autonomous Agents '01. Montreal. pp. 47-51.
- MC CRAE, R. & JOHN, O.P. (1992). An introduction to the Five-Factor model and its applications. *Journal of Personality*. Vol. 60, pp. 175-215.
- MOFFAT, D. (1997). Personality parameters and programs. In R. TRAPPL & P. PETTA, Eds. *Creating personalities for synthetic actors*. Springer LNCS 1195, pp. 120-165.
- MOON, Y. & NASS, C. (1996). How "Real"! are Computer Personalities? *Communication Research*, 23(6), pp.651-674.
- MOON, Y. & NASS, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interactions. *International Journal of Human-Computer Studies*. Vol 49, pp. 79-94.
- NASS, C., MOON, Y., FOGG, B. J., REEVES, B. & DRYER, C.D. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*. Vol.43(2), pp. 223-239.
- NASS, C., FOGG, B.J., & MOON, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*. Vol. 45(6), pp. 669-678.
- NASS, C., ISBISTER, C. & LEE, E.J.(2000). Truth is beauty: Researching embodied conversational agents. In J. CASSELL, J. SULLIVAN, S. PREVOST & E. CHURCHILL, Eds. *Embodied Conversational Agents*, MIT Press, Cambridge, Ma. pp. 374-402.
- NICHOLSON, A.E. & BRADY, J.M. (1994). Dynamic belief networks for discrete monitoring. *IEEE Transactions on Systems, Men and Cybernetics*, 24(11), pp. 1593-1610
- OATLEY, K. & JOHNSON-LAIRD, P.N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*. Vol. 13 pp. 29-50.
- ORTONY, A., CLORE, G.L. & COLLINS, A. (1988). *The cognitive structure of emotions*. Cambridge University Press, Cambridge, MA.
- ORTONY, A. (1988). Subjective importance and computational models of emotions. In V. HAMILTON, G.H. BOWER & N.H. FRJIDA, Eds. *Cognitive perspectives on emotion and motivation*. Kluwer, pp. 321-333.
- ORTONY, A.(2002) On making believable emotional agents believable. In *Emotions in humans and artifacts*. MIT Press, Cambridge, MA.

- PAIVA, A. Ed. (2000). *Affective Interactions*. Springer LNAI 1814.
- PEARL, J.(2000). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufman Publishers.
- PELACHAUD, C., BADLER, N.I. & STEEDMAN, M.(1996): Generating Facial Expressions for Speech. *Cognitive Science*, 20, 1. Pp.1-46.
- PELACHAUD, C., CAROFIGLIO, V., DE CAROLIS, B., DE ROSIS, F. & POGGI, I. (2002). Embodied Contextual Agent in Information Delivering Application. In *Proceeding of the First International Joint Conference on Autonomous Agents & Multi-Agent Systems*. Bologna, Italy.
- PELACHAUD, C. & POGGI, I. (to appear). Subtleties of Facial Expressions in Embodied Agents. *Journal of Visualization and Computer Animation*.
- PFEIFER, R. & NICHOLAS, D.W. (1985). Toward computational models of emotion. In L. STEELS & J.A. CAMPBELL, Eds. *Progress in Artificial Intelligence*. Ellis Horwood, Chichester, pp. 184-192.
- PFEIFER, R.(1988). Artificial Intelligence models of emotion. In V. HAMILTON, G.H. BOWER & N.H. FRJIDA. Eds. *Cognitive perspectives on emotion and motivation*. Kluwer, 1988, pp. 287-319.
- PICARD, R.W. (1997). *Affective Computing*. The MIT Press.
- PLUTCHIK, R. (1980). A General psychoevolutionay theory of emotion. In R. PLUTCHIK & H. KELLERMAN, Eds. *Emotion: Theory, research and experience*. Vol 1. Academic Press. New York.
- POGGI, I. & MAGNO CALDOGNETTO, E. (1996). A score for the analysis of gestures in multimodal communication. In Proceedings of the Workshop on the *Integration of Gesture and Language in Speech*. Applied Science and Engineering Laboratories. L.Messing, Newark and Wilmington, Del, pp. 235-244.
- POGGI, I. & PELACHAUD, C. (1998). Performative Faces. *Speech Communication*, 26, pp. 5-21.
- POGGI, I. & PELACHAUD, C. (2000). Emotional meaning and expression in animated faces. In A. PAIVA, Ed. *Affective Interactions*. Springer, pp. 182-195.
- POGGI, I., PELACHAUD & DE ROSIS, F. (2000). Eye communication in a conversational 3D synthetic agent. In *AI Communications, Special Issue on Behavior Planning for Life-Like Characters and Avatars*, 13, 3. Pp. 169-181.
- PRENDINGER, E. & ISHIZUKA, M. (2001): Communicative behavior of socially situated agents. In C. PELACHAUD & POGGI, I., Eds. *Proceedings of the Workshop on "Multimodal Communication and Context in Embodied Agents"*. Autonomous Agents 2001, pp. 41-46.
- RAO, A.S. & GEORGEFF, M.P. (1991). Modeling rational agents within a BDI-architecture. In J. ALLEN, R. FIKES & R. SANDEWALL, Eds. *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*. Morgan Kaufman. pp. 473-484.
- RICKEL, J. & JOHNSON, L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. *Applied Artificial Intelligence*, 13, pp. 343-382.
- ROSEMAN, I.J. (1991) Appraisal determinants of discrete emotions. *Cognition and Emotion*, 5, 3, pp. 161-200.
- SMITH, C.A. & SCOTT, H.S. (1997) A componential approach to the meaning of facial expressions. In J.A RUSSELL & J.M.FERNANDEZ-DOLS, Eds. *The psychology of facial expression*. Cambridge University Press, Cambridge, UK, pp. 229-254.
- TAKEUCHI, A. & NAGAO, K. (1993) Communicative Facial Displays as a New Conversational Modality. *Proceedings of ACM/IFIP INTERCHI'93*, Amsterdam, pp.187-193.

TAYLOR, I.C., MC INNESS, F.R., LOVE, S., FOSTER, J.C., & JACK, M.A. (1999). Providing animated characters with designated personality profiles. In S. PREVOST & E. CHURCHILL, Eds. *Proceedings of the Workshop on Embodied Conversational Characters*. Tahoe City. pp. 87-92.

TRINDIKIT website: <http://www.ling.gu.se/research/projects/trindi>

WALKER, M.A., CAHN, J.A. & WHITTAKER, S. J. (1997). Improvising linguistic style: Social and affective bases for Agent personality. *Proceedings of Autonomous Agents*, pp. 96-105.

WILKS, Y. (1999). *Machine Conversations*. Kluwer Academic Publishers.

ZUKERMAN, I & LITMAN, D. (2001). Natural language processing and user modeling: synergies and limitations. *User Modeling and User-Adapted Interaction*, 11, 1-2, 2001. Pp. 129-158.