

# Behavior Planning for a Reflexive Agent

Berardina De Carolis<sup>3</sup>, Catherine Pelachaud<sup>1</sup>, Isabella Poggi<sup>2</sup>, and Fiorella de Rosis<sup>3</sup>

<sup>1</sup>Dipartimento di Informatica e Sistemistica, Università La Sapienza, Roma

<sup>2</sup>Dipartimento di Scienze dell'Educazione, Università Roma 3

<sup>3</sup>Dipartimento di Informatica, Università di Bari

## Abstract

The aim of our research is to build a Reflexive Agent, that is able to either manifest an emotion it is feeling or to hide it. If the Agent decides to manifest its emotion, it can establish what verbal or nonverbal signals to employ in its communication and how to combine and synchronize them. In the decision of whether to express an emotion in a given context, a number of factors are considered, such as the Agent's own personality and goals, the Interlocutor's characteristics and the context. In planning how to communicate an emotion, various factors are considered as well: the available modalities (face, gaze, voice etc); the cognitive ease in producing and processing the various signals; the expressiveness of every signal in communicating specific meanings; and, finally, the appropriateness of signals to social situations.

## 1 Introduction

Artificial Agents are not yet endowed with the capacity of 'feeling' emotions, mainly because emotions would imply an involvement of the hardware of a machine and this will be impossible, or at least difficult, still for some years [Picard, 1997]. But let us suppose that we have an agent that is able to feel emotions; what could it do? How could it behave? Would it manifest its emotions to a potential interlocutor or would it ruminate on its own emotion by itself, without showing its feeling?

The effects of sharing an emotion, to the person who feels, are described in the psychology literature [Rimé, 1987]: it has been shown, for example, that people who have undergone severe shocks and share their emotions are better adapted, undergo fewer heart and pressure problems and have longer life [Pennebaker, 1989]. Of course, people differ in their proneness to display emotions: some of us are very impulsive and tend to display their emotions whatever the consequences of this

display, while others are more reflexive and follow the rule "think it over seven times seven before". In fact, if I show to my boss that I am angry at him, this may make me feel better at the moment but I may risk being fired. In some cases, though, this may also show that I'm not afraid of him, and he might respect or admire me. Now, the decision of whether to display my anger at my boss may be a very cold, calculated and utilitarian one. But we may also decide whether to display our emotion on the spot, in a not so conscious and rational way; still, also in this case we think it is possible to speak of a "decision" of whether to display emotions, even if this decision is not deliberate and conscious. The aim of our research is to build an agent that is able to express its emotions but also to refrain from expressing them: a reflexive, not impulsive agent.

The idea is to build a behavior planner that starts from a discourse plan which represents the information content and the structure of a "non-affective" text, and enriches it so as to include the 'order' of displaying some emotion. This emotion will be finally expressed by verbal or non-verbal signals, or both. This generation process allows our system to synchronize various kinds of signals and to scatter them across the media (voice, face, gesture). For example: emphasis (or topic-comment) information may be outputted by means of a pitch accent, by raising eyebrows or with a head nod. The occurrence of all signals at the same time denotes an emphasis of the emotion to be manifested. Therefore, representing detailed information on verbal and non-verbal signals in the behaviour planner ensures a more subtle and refined multi-modal discourse generation.

After reviewing works related to ours, we present a general overview of the factors that affect the decision of whether to display an emotion. Then, we give an overview of our system by emphasizing how our discourse plan is enriched with emotion display information, and we provide an example of how our system works. We end this paper with some concluding remarks and a view on our future work.

## 2 Related Work

In the construction of embodied agents capable of expressive and communicative behaviors, an important step is to reproduce affective and conversational facial expressions on synthetic faces [Ball and Breese, 2000, Cassell *et al.*, 1994, Cassell *et al.*, 1999, Lester *et al.*, 2000, Poggi *et al.*, 2000, Rickel and Johnson, 1999]. For example, REA is an interactive agent that is able to converse with a user in real-time [Cassell *et al.*, 1999]. REA exhibits refined interactional behaviors such as gestures for feedback or turn-taking functions. Cassell and Stone [Cassell and Stone, 1999] designed a multi-modal manager whose role is to supervise the distribution of behaviors across the several channels (verbal, head, hand, face, body and gaze). Cosmo [Lester *et al.*, 2000] is a pedagogical agent particularly keen on space deixis and on emotional behavior: a mapping between pedagogical speech acts and emotional behavior is created by applying Elliott's theory [Elliott, 1992]. Ball and Breese [Ball and Breese, 2000] apply bayesian networks to link emotions and personality to (verbal and non-verbal) behaviors of their agents. André *et al.* developed a rule-based system to simulate dialogues between lifelike characters with different personality traits (extroversion and agreeableness) [André *et al.*, 2000]. Marsella developed an interactive drama generator, in which the behaviors of the characters are consistent with their emotional state and individuality [Marsella, 2000].

In most of the mentioned agents, behaviors are viewed as responses to events and actions and to the way they affect emotional reactions: what is simulated is how the Agent responds with an emotion to what happens in the external context, and how an emotion affects its behavior [Marsella, 2001]. The system we propose in this paper aims at developing an Agent that is able to be affected by the emotion, but at the same time to decide whether to express it or to refrain from expressing it, not only on the basis of the emotion *per se* but also according to the context in which interaction takes place. Formalization of this reasoning considers, as we will see, several variables, such as the Agent's goals and personality and information about the Interlocutor.

## 3 Emotion Triggering Factors

Emotions are a biological feed-back device whose function is to provide information about the state of achievement or thwarting of our most important goals. Every time something relevant happens (or is assumed to happen) in the environment, in such a way that an important goal of the Agent (the goal of survival, of body safety, of reproduction etc) is, or is likely to be, achieved or thwarted, the feed-back device of Emotion is activated. This consists of a set of somatic, physiological, psychological, expressive and motivational issues that alert the Agent and, at the same

time, provide the necessary energy and resources for reaction [Castelfranchi, 2000].

Ortony *et al.* [Ortony *et al.*, 1988] and, subsequently, Elliott [Elliott, 1992] defined three entities that may be responsible for the arousal of an emotion: events, actions and objects. The occurrence of an event or some particular aspects of an object may fire appraisal of an emotion; actions of oneself or others can give rise, as well, to emotions (say, reproach or shame). Based on the Agent's interpretation of the situation, an emotion is triggered. In this work, we refer to emotion types as defined in Elliott's Affective Reasoner, in which 24 emotions are described [Elliott, 1992].

## 4 Emotion Regulation Factors

The expressive part of the emotional reaction (the fact, for example, that we open eyes wide in fear, or make a frown in anger, or blush in shame) usually has a precise function. Showing my rival I am angry with him may induce him to leave the field; showing my terror with wide open eyes may alert other co-specifics that a serious danger is present. This is why all theories of emotions also elaborate on the expressive aspects of the emotions and consider the expressive device as an integral part of the emotion itself [Ekman and Friesen, 1975].

Sometimes, though, expressing our emotion may be dangerous rather than useful. If I show fear in front of a rival, this may give him a weapon and may oblige me to leave the field. Humans, then, learned to be flexible in the use of the expressive part of the Emotion syndrome: that we are biologically endowed with a repertoire of display devices does not necessarily mean that any time we feel an emotion, we immediately and unthoughtfully display it. A famous elaboration around the topic of what drives the decision of whether to display an emotion is the notion of *display rules*, that is of culture-dependent rules that establish when, how and to whom to express one's own emotions. Ekman and Friesen studied how people tend to intensify, de-intensify, hide or mask their emotions according to social and cultural norms [Ekman and Friesen, 1975].

In this paper, we propose a formalization of some of these rules: we concentrate on whether one displays one's emotion or not without taking into account the subtlety of more intense or less intense display or, *a fortiori*, the masking of felt emotions (this will be the subject of our future work). We start from the following question. Suppose I am feeling a quite strong emotion, triggered by an event, an action or an object: on what basis will I decide whether to display it or not? Which are the factors that affect this decision? In our view, this depends on two aspects of the emotion: on one side, on the very nature of the emotion itself (emotional nature [Ortony *et al.*, 1988, Castelfranchi, 2000]); on the other side, on its interaction with 'scenario factors' [Poggi and Bartolucci, in prep.]. In Sect. 4.2 we overview some of these factors, while being conscious that the factors and

the associated values we are proposing may not be exhaustive, due to the complex nature of the emotional phenomena.

## 4.1 Emotional Nature

**Emotion Valence:** emotions may be either positive (when a goal is achieved, i.e., a desirable event occurs) or negative (when a goal is thwarted, an undesirable event occurs): feeling them may therefore be *pleasant* or *unpleasant*.

**Emotion Social Evaluation:** emotions are subject to social evaluations; for instance: showing envy is sanctioned [Castelfranchi, 2000]). They may therefore be *approved* or *sanctioned*.

**Emotion Addressee:** we may distinguish between social emotions (that are necessarily felt **towards** some person, as love or sense of guilt) and non-social emotions (as joy or fear). Social emotions may be addressed to the interlocutor itself (I) or towards a third person (O).

## 4.2 Scenario Factors

**Agent's Display Motive:** we call Display Motive the reason - the specific goal - that induces us to display a particular emotion in a particular situation [Poggi and Bartolucci, in prep.]. An Agent may want her Interlocutor(s) to feel the same emotion she is feeling. Here, the display motive is the goal that others feel empathy with her, that, at least, they participate to her emotion. In other cases the Agent displays her emotion because she wants the other to console her, to give her advice, to help her, to pity her or to reassure her about her feeling that emotion. Sometimes, she may just want an objectivation, in order to better understand herself. If the Agent's goal is just to give vent to her emotion, this goal does not imply another person: she wants to display it just to give vent to the physiological energy caused by the emotion, without needing another person to cause this process. There are, though, other cases where the Agent needs the presence of another person to display her emotion: I may confide a sad experience to a person in order to show her she's not more unlucky than others, or to strengthen relationships; even, I may show admiration to a person, to adulate him. A non-exhaustive list of the Display Motive factor includes the following: *vent, empathy, consolation, advice, help, reassurance, objectivation...*

**Agent's Personality:** an impulsive person will tend to display emotions more often than a shy or ruminative (non-impulsive) person.

**Interlocutor's Features:** the Agent may take into account, as well, some features she attributes to the Interlocutor. The following features may be relevant:

- **Interlocutor's Personality:** the Interlocutor's Personality interacts with the Agent's Display Motive: if I look for empathy, but I believe that you are a 'cold' and 'egocentric' person, I will probably avoid displaying my emotion to you. This factor also interacts with the specific emotion to display: I will not display my fear to someone who is apprehensive. A non-exhaustive list of these factors is the following: *envious, modest, egocentric, altruist...*
- **Interlocutor's Cognitive Capacity:** the factors in the following categories may take Yes/No values.
  - *Comprehension:* ability to cognitively understand the causes of the Agent's emotion.
  - *Experience:* previous experience, by the Interlocutor, of similar events that elicited the Agent's emotion. Experience interacts with the Display Motive.
  - *Problem solving:* the planning capacity of the Interlocutor to solve practical problems may be relevant (e.g. when the Agent's Display Motive is to get help or advice).

**Agent-Interlocutor Role Relationship,** or power relationship between people: I may avoid displaying my anger to my boss because he has power over me and can retaliate. For the sake of simplicity, we consider the following power relationships between agent (Ag) and interlocutor (I): *Ag has power over I, I has power over Ag, neutral*.

**Agent-Interlocutor Personal Relationship,** that is, the aggressive or friendly attitude between them: this may take four values, depending on the sense of the relationship (either Ag-to-I or I-to-Ag) and on its valence (either aggressive or adoptive). Notice that we call 'adoptive' a relationship in which one is taking care of the other's goals. The values are then: *Ag adoptive I, Ag aggressive I; I adoptive Ag, I aggressive Ag*.

**Type of social interaction:** the last important factor in the regulation of emotions is whether interaction occurs in public or not. This interacts with the Agent's Personality: the more impulsive I am, the more I'll be likely to display emotions even in public; the more shy I am, the more interaction in public will inhibit my display.

## 5 The Agent's Behavior Planner

In order to build the reflexive component of our Agent's mind and to interface it with its body, we developed a prototype of a Behavior Planner that is based on the architecture shown in Figure 1. We will use the following example to describe, step by step, how the system works.

*"A 3 year old girl wants to give a present to her mummy; she cuts some stripes from mummy's suit and knots them together as to make a ribbon. Mummy comes in and sees the massacre".*

In such a situation, the Agent (Mummy) reproaches her daughter of such an action. The input of our Discourse Generator is therefore the communicative goal: Reproach (Mummy Daughter 'cut suit'). The first step of the process consists in producing a 'discourse plan' (D-Plan) to achieve this communicative goal. This plan does not yet contain any emotional content or reference (we call it a "non-affective" D-Plan); it can be produced by a planning algorithm, from a *library of plan operators* [Moore and Paris, 1993] or may be retrieved from a *library of plans* [De Carolis *et al.*, 2000].

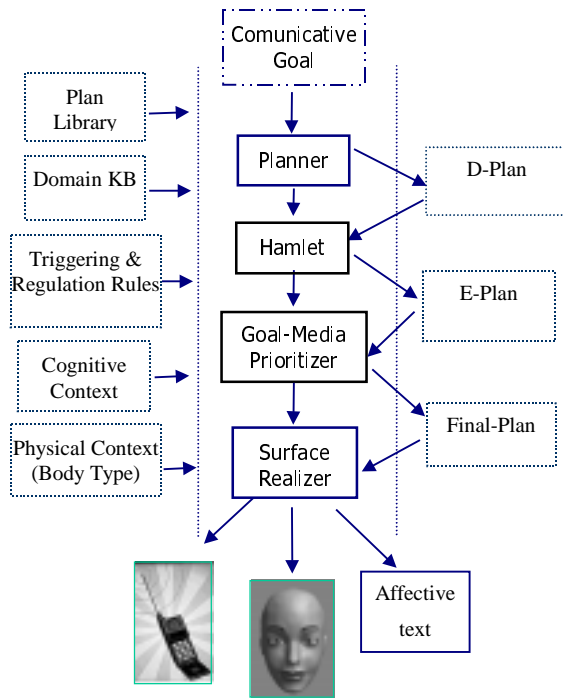


Figure 1. The Behavior Planner

In both cases, the output is a plan tree that defines the content and order of presentation of information to be conveyed in the discourse. In our case, the planner uses a library of non-instantiated plans to select the one that satisfies the posted goal. If such a plan is not available in the library, the planner builds it by appropriately combining the existing plans. The generic plan is, then, instantiated by filling the variable-slots with domain data. Part of the D-plan that is generated in our example is shown in Figure 2.

After generating the "non-affective" D-Plan, the *Hamlet* Component is activated. *Hamlet* takes as an input the *triggering rules*, that decide whether to fire the emotion; it then applies the *regulation rules* by processing the Scenario Factors, to decide whether to display the fired emotion. The triggering rules are drawn by Elliott's Affective Reasoner [Elliott, 1992], while the

regulation rules are similar to the *display rules* in [Ekman and Friesen, 1975]; they are, however, enriched to encompass the Scenario Factors described above. As we anticipated in Section 4, our System does not take only cultural norms into account, but a very detailed model of the Context and of the Interlocutor. The structure of the Hamlet module is shown in Figure 3.

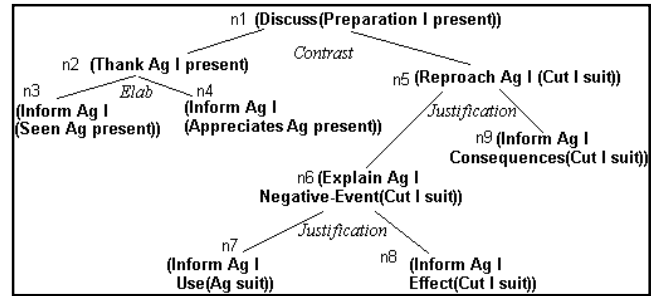


Figure 2: An example of D-Plan

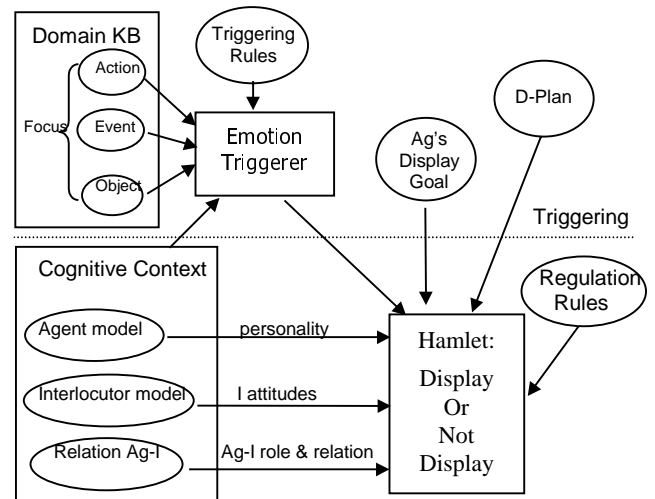


Figure 3: The Hamlet Module

Let us now see some examples of *triggering* and *regulation rules*.

The general structure of the triggering rules is the following:

IF DC-Cond THEN (Feel Ag e),

where DC-Cond represents a condition on the context or the domain, Ag denotes the agent that is conveying the signal, and e an emotion. The evaluation of DC-Cond as, for example, a Desirable or Undesirable Event or Object, is determined through a pattern-matching with the Domain KB; Feel is a predicate which applies to the two mentioned terms, to denote that 'the agent Ag feels the emotion e'. For social emotions, this predicate applies, instead, to three terms: the Agent, the emotion,

and the one to whom the emotion is addressed (the Interlocutor I or another agent O). Events, Objects and Actions are represented, in D-Plan, in the ‘discourse focus’ that is associated with every node.

Let, for instance,  $n_5$  be the considered node in D-Plan; the focus of this node is (Cut I suit), which is an ‘undesirable action’, to the agent Ag. The triggering rule that will be applied in this case is the following:

- **R1:** IF (Focus( $n_i$ ) act) AND (Undesirable act)  
AND(Performed I act) AND (Blames Ag I act))  
THEN (Feel Ag Anger)

That is: “Ag feels anger if the focus of  $n_i$  refers to an ‘undesirable’ action that was performed by the interlocutor I and Ag disapproves this action.

In our example, **R1** is activated because the Domain KB mentions that the action ‘cut suit’ is labelled as ‘Undesirable’ in I is the agent who performed it.

Once an emotion has been triggered, the *regulation rules* are activated, to decide whether this emotion has to be displayed in the given context. This decision is taken by considering context features, such as: the Agent's display motive, personality and role, the Interlocutor's attitudes and the relation between the two Agents.

The general structure of the *regulation rules* is the following:

```
IF (Feel Ag e) AND DC-Cond
  THEN (Display Ag e),
```

or

```
IF (Feel Ag e) AND DC-Cond
  THEN NOT (Display Ag e).
```

As for the *triggering rules*, DC-Cond is instantiated, also in this case, by matching its variables with the content of the Domain KB. If we go back to our example, the following *regulation rules* are activated to decide whether to display the emotion:

- in the first case, Mummy is angry but she loves her daughter and knows she could not understand how destructive was the action she was doing. Therefore, Mummy will not display her anger.

```
R2: IF (Feel Ag Anger) AND (Adoptive Ag I)
  AND NOT (UnderstandConsequences I act)
  THEN NOT(Display Ag Anger)
```

- in the second case, Mummy thinks her daughter had comprehension capacity:

```
R3: IF (Feel Ag Anger) AND (Adoptive Ag I)
  AND (UnderstandConsequences I act)
  THEN (Display Ag Anger).
```

The enriched plan E-Plan is produced by augmenting D-Plan with the ‘display’ commands through application of these rules. The E-Plan corresponding to the second case is shown in Figure 4. Notice that every node inherits the ‘Display’ label from its parent-node, while the ‘NotDisplay’ label is simply not inserted in the plan:

so, anger will be felt but not displayed in this case, while joy is felt and displayed, because ‘present’ is a ‘desirable object’ and there is no reason for hiding the joy emotion it triggers. On the contrary, anger would be felt *and* displayed in the second case.

In the next step, E-Plan is transformed into an XML document and is sent to the Goal-Media prioritizer. This module decides how to distribute the signals that enable displaying an emotion over the different media. For example: Anger might be displayed in one of the following ways, or, if very intense, in the three of them: verbally (through the hyperbole “*I told you a hundred times...*”); vocally (through a loud or a high pitch voice), and facially (with tense lips and an angry frown). So, this module revises the enriched plan by deciding the type of non-verbal signals to employ at every conversational step, their combination and their synchronization with verbal communication. This is done according, again, to the context and also to the type of body that has been chosen for the Agent. The XML-output of this module is the final presentation plan, where the media have been instantiated and, when necessary, synchronized or linearized. When the Goal-Media prioritizer has selected the appropriate output form (2D/3D model, audio, text...), the XML tags are interpreted and translated into parameters to drive the given output signals.

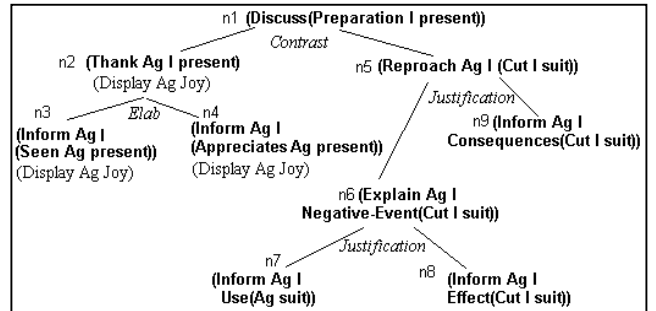


Figure 4. An example of E-Plan

At present, we are concentrating on the animation of a 3D facial model: the XML tags are converted into MPEG-4 facial parameters that drive the final animation. The XML tags are interpreted and translated into parameters to drive the given output signals; the natural language ‘surface’ generation step is adapted, as well, to the emotional state of the Agent. To go back to our example, in the first case, the subplan originating from node  $n_5$  will be rendered by the following sentence, pronounced with a ‘not angry’ face expression:

*"The ribbon you used is made from my suit, that I need for my work. If you cut it, I can't wear it anymore."*

While in the second case, the Mummy will say, with an ‘angry’ face:

"The ribbon you used is made from my suit, that I need for work. I already told you a hundred times not to touch to my things!"

## 6 Conclusions and Future Work

In this work, we have presented the architecture of a system that builds a reflexive agent, that is, an agent who, when feeling an emotion, "decides" whether to display it immediately or not. We have also defined the different elements on which this "decision" is based (emotional nature and scenario factors). Our behaviour planner formalises these elements and then plans how to produce and synchronize the verbal and nonverbal parts of a discourse that satisfy a given communicative goal. Two sets of rules are considered: triggering rules, that fire an emotion, and regulation rules, that make the agent 'reflexive'. In the future, we plan to evaluate how believable is our Reflexive Agent.

## References

- [Andre *et al.*, 2000] E. Andre, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogues for animated presentation teams. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. Cambridge, MA: MIT Press, 2000.
- [Ball and Breese, 2000] G. Ball and J. Breese. Emotion and personality in a conversational agent. In S. Prevost J. Cassell, J. Sullivan and E. Churchill editors. *Embodied Conversational Characters*. Cambridge, MA: MIT Press, 2000.
- [Cassell *et al.*, 1994] J. Cassell, C. Pelachaud, N.I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics Proceedings, Annual Conference Series*, pages 413--420. ACM SIGGRAPH, 1994.
- [Cassell *et al.*, 1999] J. Cassell, J. Bickmore, M. Billinghamurst, L. Campbell, K. Chang, H. Vilhjalmsen, and H. Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of CHI'99*, pages 520--527, Pittsburgh, PA, 1999.
- [Cassell and Stone, 1999] J. Cassell and M. Stone. Living hand and mouth. Psychological theories about speech and gestures in interactive dialogue systems. In *Proceedings of AAAI99 Fall Symposium on Psychological Models of Communication in Collaborative Systems*, 1999.
- [Castelfranchi, 2000] C.Castelfranchi. Affective Appraisal versus Cognitive Evaluation in Social Emotions and Interactions. In A.Paiva editor. *Affective Interactions. Towards a New Generation of Computer Interfaces*. Berlin: Springer, 2000.
- [De Carolis *et al.*, 2000] B. De Carolis., C. Pelachaud and I. Poggi. Verbal and nonverbal discourse planning. In *Proceedings of International Agents 2000 Workshop on Achieving Human-Like Behavior in Interactive Animated Agents*. Barcelona, 2000.
- [Ekman and Friesen, 1975] P.Ekman and W. Friesen. *Unmasking the Face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., 1975.
- [Elliott, 1992] C. Elliott. An Affective Reasoner: A process model of emotions in a multiagent system, Technical Report No. 32 of The Institute for the Learning Sciences Northwestern University, 1992.
- [Lester *et al.*, 2000] J.C. Lester, S.G. Stuart, C.B. Callaway, J.L. Voerman, and P.J.Fitzgerald. Deictic and emotive communication in animated pedagogical agents. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, 2000.
- [Mann and Thompson, 1988] W.C Mann and S. Thompson. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text*, 8(3):243-281, 1988.
- [Marsella, 2000] S C Marsella. Sympathy for the Agent: Controlling an Agent's nonverbal repertoire. In: Achieving Human-like Behavior in Interactive Animated Agents. 4<sup>th</sup> Int Conf Autonomous Agents, Barcelona, 2000.
- [Moore and Paris, 1993] J.D. Moore and C. Paris. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*, 19(4), 651-694, 1993.
- [Ortony *et al.*, 1988] A. Ortony, G.L. Clore and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
- [Pennebaker, 1989] J.W. Pennebaker. Confession, inhibition and disease. in L.Berkowitz (Ed.), *Advances in Experimental Social Psychology* 22. Orlando, FL.: Academic Press, 1989, 211-244.
- [Picard, 1997] R. Picard. *Affective Computing*, MIT Press, 1997.
- [Poggi *et al.*, 2000] I. Poggi, C. Pelachaud, and F. de Rosis. Eye communication in a conversational 3D synthetic agent. *Special Issue on Behavior Planning for Life-Like Characters and Avatars of AI Communications*, 2000.
- [Poggi and Bartolucci, in prep.] I.Poggi and L.Bartolucci. Factors affecting the displaying of Emotions. In preparation.
- [Rickel and Johnson, 1999] J. Rickel and W.L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343--382, 1999.
- [Rimé, 1987] B.Rimé. Le partage social des Emotions. In B.Rimé et K. Scherer editors. *Les Emotions*. Genève: Delachaux et Niestlé, 1987.