Cognitive Emotion Modeling in Natural Language Communication [1]

Valeria Carofiglio, Fiorella de Rosis, Nicole Novielli

Department of Informatics,

Via Orabona 4

70124 Bari, Italy

{carofiglio,derosis,novielli}@di.uniba.it

Abstract. This paper describes some psychological theories that are at the foundation of research on cognitive models of emotions, to then review the most significant projects in this domain, in the recent years. The review is focused on probabilistic dynamic models, due to the key role of uncertainty in the relationships among the variables involved: the authors' experience in this domain is discussed, by outlining problems still open. Two aspects are discussed in particular: how probabilistic emotion models can be validated and how the problem of emotional-cognitive inconsistency can be dealt with in probabilistic terms.

Keywords: cognitive emotion models, dynamic belief networks, cognition-emotion inconsistency

1 Introduction

Computer science recently began, with success, to endow natural language dialogues with emotions, by mainly relying on the OCC theory. More recent works went beyond this classification, to consider categories of emotions that occur frequently in human-computer communication (Carberry and de Rosis, 2008). This Chapter describes how emotion activation and interpretation theories may contribute to build affective state representations of human and artificial agents, and which are the problems in building and validating these models.

---

[1] *In "Affective Information Processing",*
J. Tao and T.Tan (Eds), Springer, in press

Section 2 will be focused on describing the affective states which may influence natural language communication. Sections 3 will briefly review some psychological theories of emotions: as appraisal theories are best suited to cognitive modeling, essential aspects of these theories will be considered in particular. In section 4, the main requirements of cognitive models that combine rational with emotional mental state components are discussed. Previous experiences in this area are analyzed briefly in Section 5, with special focus on models considering uncertainty and time-dependency in emotion activation and decay. In the same section, the authors' own formalism for emotions modeling is then introduced and compared with these models. A general method for validating these probabilistic models is outlined in Section 6, with an example of application. Finally, Section 7 reflects critically on the problem of cognitive-emotional inconsistency and how this may be dealt with in probabilistic models.

2 Affective States Relevant in NL Communication

The term "affective state" includes several aspects, that vary in their degree of stability and in their "object-oriented specificity". In this contribution, the following aspects will be considered in particular (Cowie, 2006):

– *Personality traits*: stable, dynamic and organized set of characteristics of a person that uniquely influences his or her cognitions, motivations, and behaviors in various situations.

– *Interpersonal stance*: a relatively short-lived state where feeling towards another person comes to the fore and inclines you to behave in a particular way towards him or her.

– *Emotions*: typically complex, episodic, dynamic and structured events. They are complex in that they typically involve many different elements: perceptions, thoughts, feelings of various kinds, bodily changes, dispositions to experience further emotional episodes. They are episodic and dynamic in that, over time, the elements can come and go, depending on all sorts of factors. They are structured in that they constitute part of an unfolding sequence of actions and events, thoughts and feelings, in which they are embedded.

3 Theories Behind Cognitive Emotion Models: Which Ones Best Fit?

Psychologists worked at decoding emotions for decades, by focusing on two main questions: (i) how can emotions be classified?; (ii) which is their functioning?, i.e. how are they triggered? How do they affect behavior? Which is the role played by cognition? Two points of view prevailed: the first one assumes that a limited set of basic emotions exists, while the second one consider emotions as a continuous function of one or more dimensions: see, e.g., the ”circumplex model” of affect (Russell, 2003). Theories following the ”discrete trend” agree on the idea that a limited set of basic emotions exists, although consensus about the nature and the number of these basic emotions has not been reached. Ekman defines a basic emotion as having specific feelings, universal signals and corresponding physiological changes (Ekman, 1999); emotions may be triggered by either an automatic (very quick) or extended (slow, deliberate and conscious) appraisal mechanism. Plutchik defines discrete emotions as corresponding to specific adaptive processes: reproduction, safety, etc (Plutchik, 1980). In Izard’s ”Differential Emotion Theory”, distinct discrete emotions are triggered by neural activation, which induces a specific experience and influences behavior and cognition (Izard, 1993). Lazarus describes nine negative (Anger, Anxiety, Guilt, Shame, Sadness, Envy, Jealousy, Disgust) and six positive (Joy, Pride, Love, Relief, Hope, Compassion) emotions, with their appraisal patterns: positive emotions are triggered if the situation is congruent with one of the individual’s goals; otherwise, negative emotions are triggered (Lazarus, 1991).

Some physiological theories assume that stimuli produce a physiological activation that creates emotions, with no cognitive intervention (James, 1984). Evolutionary theories assume that emotions were inherited during evolution and are automatically triggered with no cognitive intervention (Ekman, 1999). On the contrary, cognitive theories assume that cognition is essential in the triggering of emotion.

Among the cognitive theories of emotions, ”appraisal theories” assume that emotions are triggered by the evaluation of a stimulus, using different criteria. To Arnold, ”appraisal” is the process for determining the significance of a situation for an individual, i.e. if a given situation is good or bad for him or herself (Arnold, 1960). Appraisal triggers an emotion that induces an ”action tendency” of attraction or repulsion, with a corresponding physiological change (to favor the individual’s adaptation to the environment). Ortony, Clore and Collins focus their theory on the cognitive elicitors of emotions. They postulate that emotions represent valenced reactions to perceptions of the world. Thus, one can be pleased or not about the consequences of an event; one can endorse or reject the actions of an agent or can like or dislike aspects of an object (Ortony et al., 1988). Lazarus assumes that the human emotional process is made up of two separated processes: appraisal (which characterizes the persons relationship with their environment), and coping (which suggests strategies for altering or maintaining this relationship). Cognition informs both these processes; it informs appraisal by building mental representations of how events relate to internal dispositions such as beliefs and goals, and informs coping by suggesting and exploring strategies for altering or maintaining the person-environment relationship (Lazarus, 1991).

These theories show several advantages for cognitive modeling:

– *A fine-grained differentiation among emotions*: Cognitive appraisal theories assume that emotions result from the evaluation of stimuli. Thus each distinct emotion corresponds to a particular appraisal pattern, that is a particular combination of appraisal criteria.

– *A differentiation (at various levels) among individuals in felt emotions*: in the same situation, several individuals can feel different emotions. Different evaluations of the situation represent cognitive mediation between stimulus and felt emotion. These may be linked to individual personality traits or to mental representations of how events relate to internal dispositions such as beliefs and goals.
– *A temporal differentiation of felt emotions by an individual*: the same individual can appraise differently the same situation at different moments, and thus feel different emotions, depending on the context.

3.1 Relationship Between Emotions and Goals

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 (Oatley and Johnson-Laird, 1987). In Lazarus's primary appraisal, the relevance of a given situation to the individual's relevant goals is assessed (Lazarus, 1991). At the same time, emotions activate goals and plans that are functional to re-establishing or preserving the well being of the individual, challenged by the events that produced them (secondary appraisal, (Lazarus, 1991)). There is a strong relationship, then, 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-level 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 (Poggi, 2005). As personality traits may be viewed in terms of weights people put on different goals, the strength of the relationship between goals and emotions also depends on personality traits. These considerations suggest the following criteria for categorizing emotions (Poggi, 2005):

– *The goals the emotion monitors*. For instance: *'Preserving self from -immediate or future- bad'* may activate distress and fear; *'achieving the -immediate or future- good of self'*, may activate joy and hope. *'Dominating others'* may activate envy. "*Acquiring knowledge and competence*" activates the cognitive emotion of curiosity. *Altruistic emotions* like guilt or compassion involve the goal of "defending, protecting, helping others". *Image emotions* like gratification involve the goal of "being evaluated positively by others"; *Self-image emotions* like pride involve the goal of "evaluating oneself positively"... and so on.

– *The level of importance of the monitored goal:* this may be linked to the type of goal but is also an expression of some personality traits (e.g. neuroticism)

– *The probability of threatening/achieving the monitored goal*: some emotions (joy, sadness) are felt when the achievement or thwarting is certain; others (hope, fear), when this is only likely;

– *The time in which the monitored goal is threatened/achieved*: some emotions (joy, sadness) are felt only after goal achievement or thwarting, others (enthusiasm) during or before goal pursuit.

– *The relevance of the monitored goal with respect to the situation*: if a situation is not relevant to any individual's goal, then it cannot trigger any emotion in the individual.

– *The relationship between the monitored goal and the situation in which an individual feels an emotion*: this is linked to the emotion valence, but also to its intensity. Some emotions belong to the same "family" but differ for their intensity (disappointment, annoyance, anger, fury).

Once an emotion is felt by an individual, an evaluation of the actions he or she may perform and their envisaged effects on the situation occurs. Coping potential is a measure of how individuals can change their goals, to restore a good relationship with the environment (Lazarus 1991).

3.2. Relationship Between Emotions and Beliefs

Emotions are biologically adaptive mechanisms that result from evaluation of one's own relationship with the environment. However, not all appraisal variables are translated into beliefs: some of them influence emotion activation without passing through cognitive processing, while others do it. These cognitive appraisals are beliefs about the importance of the event, its expectedness, the responsible agent, the degree to which the event can be controlled, its causes and its likely consequences: in particular (as said) their effect on goal achievement or threatening. Different versions of this assumption can be found in various cognitive appraisal theories of emotion (Arnold, 1960; Lazarus, 1991; Ortony et al., 1988, Scherer, 2004). In individual emotions, cognitive appraisals are first-order beliefs while in social emotions second-order beliefs occur as well.

Beliefs therefore influence activation of emotions and are influenced by emotions in their turn (Frijda and Mesquita, 2000). They may trigger emotions either directly (as emotion antecedents) or through the activation of a goal (Miceli et al, 2006). On the other hand, psychologists agree in claiming that emotions may give rise to new beliefs or may strengthen or weaken existing beliefs. Classical examples are jealousy, which strengthens perception of malicious behaviors, or fear, which tends to increase perception of the probability or the amount of danger of the feared event (this issue in Section 7). As events that elicit emotions may fix beliefs at the same time, this effect may strengthen the cyclical process of belief holding - emotion activation - belief revision: when you hear at the TV about a plane that fell down, your fear strengthens your beliefs about the risks of flying. These beliefs may be temporary, and last for as long as the emotion lasts. But they may become more permanent, as a result of a 'rumination and amplification' effect: in this case, for instance, feeling fearful after looking at the TV service may bring one to detect more fearful stimuli in the news, which increase fear in their turn.

Finally, emotions may influence beliefs indirectly, by influencing thinking: they are presumed to play the role of making human behavior more effective, by activating goals and orienting thinking towards finding a quick solution to achieving them. They hence influence information selection by being biased towards beliefs that support emotional aims.

4 How to Apply These Theories to Build Emotion Models?

The lesson from the psychological theories mentioned so far is that, in trying to formalize a computationally tractable model of emotion, researchers have to deal with a very complex reality. Individuals can experience several emotions at the same time, each with a different intensity, which is due to the importance of the goal, the likelihood of the impact of the perceived event on that goal, the level of surprise about the event, the measure of how probable it is that the event will occur, the level of involvement in the situation that leads to emotion triggering. Once activated, emotions can decay over time with a trend which depends on the emotion and its intensity; duration is also influenced by personality traits and social aspects of the interaction. Hence, the following factors may be listed as aspects to necessarily include in a computational model of emotions:

a. Appraisal Components of Emotions

How is it that two persons report feeling different emotions in the same situation? Clearly, the main source of difference is due to the different structure of beliefs and goals of the two individuals: the goals they want to achieve, the weights they assign to achieving them and the structure of links between beliefs and goals. Differences among experienced emotions may be due to the link between situation characteristics (e.g. which event – together with its consequences – which social aspects – role of individuals involved in the situation, their relationship – which individual’s personality traits) and emotion components (appraisal variables but also other internal dispositions – beliefs e goals). They may be due, as well, to the mutual links between emotion components themselves (van Reekum and Scherer, 1997).

b. Influence of Personality Factors

Individual differences can alter experienced emotions. 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 (Mc Crae and John, 1992): for instance, neuroticism. These traits imply, in a sense, a lower threshold in emotion feeling (Ortony et al., 1988). For instance, a 'shy' person is keener to feel 'shame', especially in front of unknown people. 'Proud' persons attribute a high weight to their goal of self-esteem, in particular to the subgoal of not depending on other people. So, every time one of these goals is achieved, the proud person will feel the emotion of pride; 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. A personality trait (proud) is therefore 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.

c. Emotion Intensity and Decay/Duration

In the OCC theory (Ortony et al., 1988), desirability, praiseworthiness and appealingness are key appraisal variables that affect the emotional reaction to a given situation, as well as their intensity: they are therefore called *local intensity variables*. For example, the intensity of prospect-based and confirmation emotions is affected by: the *likelihood* that the prospected event will happen; the *effort* (degree of utilization of resources to make the prospected event happen or to prevent it from happening) and the *realization* (degree to which the prospected event actually happens). The intensity of fortune-of-others emotions is affected by: *desirability* for the other (how much the event is presumed to be desirable for the other agent), *liking* (has the individual appraising the situation a positive or a negative attitude toward the other?); *deservingness* (how much the individual appraising the situation believes that the other individual deserved what happened to him or her). Some *global intensity variables* also affect the intensity of emotions: *sense of reality* measures how much the emotion-inducing situation is real; *proximity* measures how much the individual feels psychologically close to the emotion-inducing situation; *unexpectedness* measures how much the individual is surprised by the situation; *arousal* measures how much the individual was excited before the stimulus.

Some of the systems which follow the OCC theory define a measure for emotion intensity in terms of some mathematical function combining the previously mentioned variables. In Prendinger's "emotion-elicitation rules" (Prendinger et al., 2002) some of these variables are selected, and emotions intensities depend on the logarithm of the sum of their exponentials. The intensity of *joy*, for instance, 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. In Elliot's emotion eliciting condition (Elliott and Siegle, 1993), the intensity of emotions also depends on a combination of some variables, in a set of twenty-two possible ones (e.g. importance, surprisingness, temporal proximity etc). A limit of this method may be in the combination of several heterogeneous measures in a single function: when different scales are employed, the effect of their combination cannot be foreseen precisely. In addition, the range of variability of the function may go outside the range of the variable that this function should measure, with the need of normalizing the obtained value or of cutting off values out of this range.

A second approach adopts a probabilistic representation of the relationships among the variables involved in emotion triggering and display. These are usually dynamic models that represent the transition through various emotional states during time. Intensity of emotions are calculated by applying utility theory (see, e.g., Conati,2002; de Rosis et al., 2003). This approach will be further analyzed in Section 5,

d. Emotion Mixing

In her book on affective computing, Rosalind Picard proposes two metaphors for describing the way emotions may mix up: in a '*microwave oven'* metaphor, juxtaposition of two or more states may occur even if, at any time instant, the individual experiences only one emotion. 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, 1997). 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. Picard evokes the *generative mechanism* as the key factor for distinguishing between emotions that may *coexist* (by mixing according to a 'tub of water' metaphor) and emotions that *switch* from each other over time (by mixing according to the 'microwave oven' metaphor). She suggests that co-existence may be due, first of all, to differences in these generative mechanisms. But it may be due, as well, to differences in time decay among emotions that were generated by the same mechanism at two distinct time instants: for instance, 'primary' emotions, like *fear*, and cognitive ones, like *anticipation*.

e. Role of Uncertainty

To represent triggering of emotions and also the two ways in which emotions may mix, the modeling formalism adopted should be able to represent their generative mechanism, the intensity with which they are triggered and the way this intensity decays over time. Logical models are probably inappropriate to handle all the parameters mentioned above. Due to the presence of various sources of uncertainty, (static/dynamic) *belief networks* (BNs/DBN) or *dynamic decision networks* (DDNs) are probably more appropriate to achieve these goals. A BN is a graphical model whose nodes represent discrete random variables and whose arcs represent dependencies between variables in terms of conditional probabilities (Pearl 1988). DBNs are an extension of BNs, to enable modeling temporal environments (Nicholson and Brady, 1994). A DBN is made up of interconnected time slices of usually structurally identical static BNs; previous time slices are used to predict or estimate nodes' state in the current time slice. The conditional probability distributions of nodes with parents in different time slices are defined by a state evolution model, which describes how the system evolves over time. Finally, DDNs (Russel and Norvig, 1995) extend DBNs to provide a mechanism for making rational decisions by combining probability and utility theory, within changing environments: in addition to nodes representing discrete random variables over the time, the networks contain utility and decision nodes.

In the following, the power of DBNs in representing the dynamic arousal, evolution and disappearing of emotions is discussed, together with the way these phenomena are influenced by personality factors and social context. The logic behind the DBN-based modeling method is illustrated in a paper in which the advantages the method offers in driving the affective behavior of an Embodied Conversational Agent are also discussed (de Rosis *et al*., 2003). Due to the importance of uncertainty in emotion activation and interpretation, the next section will be focused on some remarkable probabilistic emotion models.

5 Some Remarkable Experiences in Emotion Modeling

In a first attempt (Ball and Breese, 2000), a simple model was proposed: emotional state and personality traits were characterized by discrete values along a small number of dimensions (valence & arousal, dominance & friendliness). These internal states were treated as unobservable variables in a Bayesian network model. Model dependencies were established according to experimentally demonstrated causal relations among these unobservable variables and observable quantities (expressions of emotion and personality) such as word choice, facial expression, speech etc. EM (Reilly 1996) simulated the emotion decay over time for a specific set of emotions, according to the goal that generated them. A specific intensity threshold was defined for each emotion to be triggered. Affective Reasoner (Elliott and Siegle, 1993) was a typical multi-agent model.

Each agent used a representation of both itself and the interlocutor's mind, evaluated events according to the OCC theory and simulated a social interactive behavior by using its knowledge to infer the other's mental state. Emile (Gratch and Marsella, 2001) extended Affective Reasoner by defining emotion triggering in terms of plans representation: the intensity of emotions was strictly related to the probability of a plan to be executed, which was responsible for the agents' goal achievement. Conati proposed a model based on Dynamic Decision Networks to represent the emotional state induced by educational games and how these states are displayed (Conati 2002). The emotions represented in this model (reproach, shame and joy) were a subset of the OCC classification; some personality traits were assumed to affect the student's goals. The grain size of knowledge representation in this model was not very fine: among the various attitudes that may influence emotion activation (first and second order beliefs and goals, values etc), only goals were considered in the model. Subsequently, the same group (Conati and McLaren, 2005) described how they refined their model by adding new emotions and learning parameters from a dataset collected from real users. In Prendinger and Ishizuka (2005), an artificial agent empathizes with a relaxed, joyful or frustrated user by interpreting physiological signals with the aid of probabilistic decision networks: these networks include representation of events, agent's choices and utility functions.

In the next sections, the authors briefly describe their work on Bayesian affective models. This was oriented in two directions: in a first model (*Emotional-Mind*), they advocated for a fine-grained cognitive structure in which the appraisal of what happens in the environment was represented in terms of its effects on the agent's system of beliefs and goals, to activate one or more *individual emotions* in the OCC classification (Carofiglio et al, in press). They subsequently (with *Social-Mind*) worked at representing how social emotions (in particular, the social attitude of users towards an Embodied Advice-Giving Agent) can be progressively recognized from the users' linguistic behavior (de Rosis et al, 2007).

a. Emotional-Mind

As anticipated, the starting point of the model is that emotions are activated by the belief that a particular important goal may be achieved or threatened (Carofiglio and de Rosis, 2005). So, the simulation is focused on the change in the belief about the achievement (or threatening) of the goals of an agent A, over time. In this monitoring system, the cognitive state of A is modeled at the time instants T1, T2,… Ti,… . Events occurring in the time interval [Ti,Ti+1] are observed to construct a probabilistic model of the agent's mental state at time Ti+1, with the emotions that are eventually triggered by these events. The intensity of emotions depends on two parameters: (1) the uncertainty in the agent's beliefs about the world and, in particular, about the possibility that some important goal will be achieved or threatened, and (2) the utility assigned to this goal.

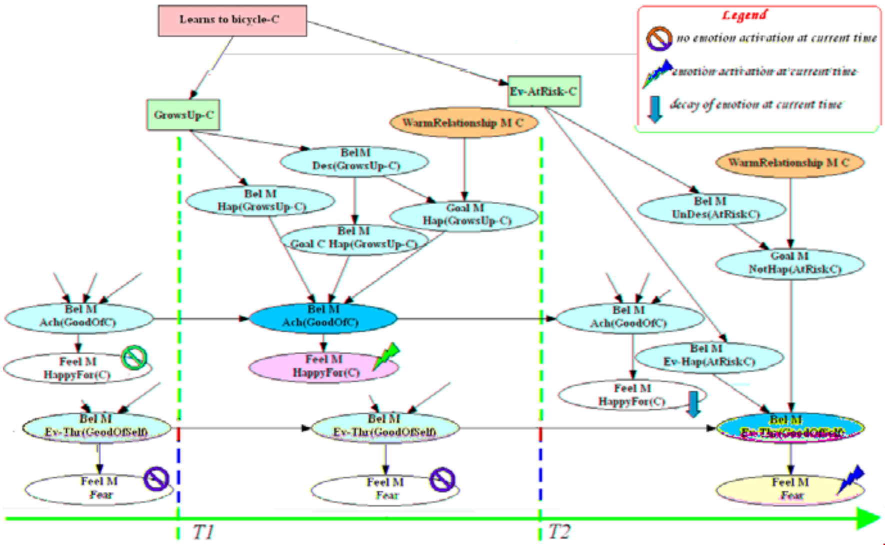

Fig.1. An example application of Emotional-Mind to activation of happy-for and fear according to the 'microwave oven' metaphor. Square boxes represent an event and its consequences in the intervals [T0,T1] and [T1,T2].Circular boxes represent appraisal variables, internal dispositions (beliefs, goals) and emotions at current time. At every time slice, triggered emotions with the associated goals are highlighted (respectively, in light blue and light green)

Figure 1 shows an example of application of this model. Here, M (a mother) is observing her little child C who just learnt to bicycle. Happy-for and fear may be activated at consecutive time slices (T1 and T2), because of positive and negative consequences of this event occurring or just imagined by M, in two distinct time intervals [T0, T1] and [T1,T2]: the child is growing up (`GrowsUp-C`) and might be at risk in a more or less near future (`Ev-AtRisk-C`).

Let us first consider the positive consequence of C learning to bicycle: as M is in a warm relationship with her child (`WarmRelationship M C`), happy-for will be triggered after believing that C is growing up (`Bel M Hap(GrowsUp-C)`), that this is a desirable event to C (`Bel M Des(GrowsUp-C)`) and hence that this is what M desires (`Goal M Hap(GrowsUp-C)`) and what M believes that C desires as well (`Bel M Goal C Happ(GrowsUp-C)`). In this situation, the goal of 'getting the good of others' is achieved (`Bel M Ach(GoodOfC)`). The intensity of this emotion depends on how much this probability varies when evidence about the mentioned desirable event is propagated in the network. It depends, as well, on the weight M attaches to achieving that goal which is, in its turn, a function of the agent's personality.

Let us now consider the negative consequence of the same event. The figure shows that fear may also be activated in M by the belief that C might fall down by cycling because of his low experience: (`Bel M Ev-Hap(AtRisk-C)`) and (`Bel M UnDes(AtRisk-C)`). In this example, the affective state of M switches among the two emotions in time.

b. Social-Mind

In this model, DBNs are employed to represent how the social attitude of users towards a human or artificial agent evolves during dialogic interaction. The definition of social attitude applied is based on Andersen and Guerrero's (1998) concept of *interpersonal warmth* as *"the pleasant, contented, intimate feeling that occurs during positive interactions with friends, family, colleagues and romantic partners...*". The main idea is to combine linguistic and acoustic features of the user move at time T with information about the context in which the move was uttered at T-1, to dynamically build a image of the user's social attitude towards the agent, as a function of its value at time T-1 and of the characteristics of the user move at time T.

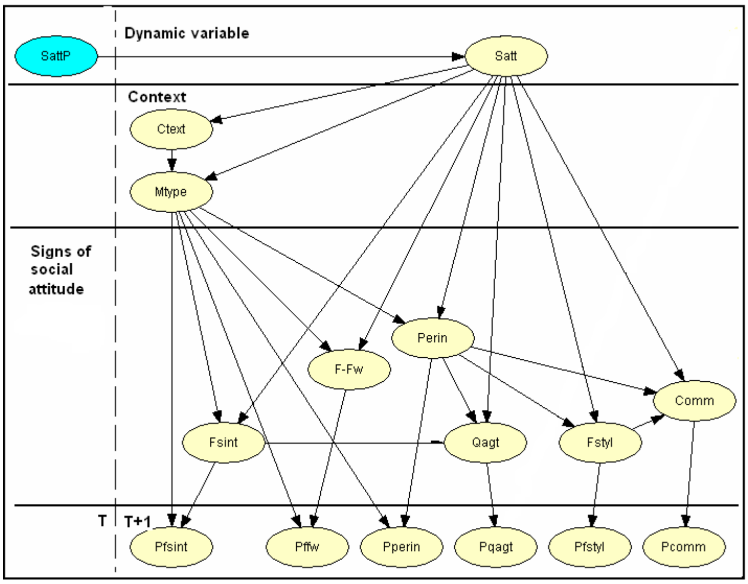

Fig. 2. Social-Mind: a generic time slice

Rather than focusing on individual emotions, *Social-Mind* is conceived to model interpersonal stances, that is long-lasting and gradually evolving affective states. In particular, the model allows representing how the 'social attitude' of users towards the application smoothly changes during interaction, by collecting evidence about a set of '*signs*' displayed by the user and conveyed by language features. Signs can be detected with a Bayesian classifier based on spotting of domain-relevant linguistic features or general semantic categories of salient (sequences of) words. The categories are defined according to psycholinguistic theories (Andersen and Guerrero, 1998; Polhemus et al., 2001), and the Bayesian classification enables computing, for every string, a value of a-posteriori probability for every sign of social attitude. Figure 2 shows the version of the model in which the observable variables (leaf nodes of the oriented graph) are the semantic categories that characterize the presence of a sign of 'warm' attitude: *personal address and acknowledgement*, *feeling*, *paralanguage*, *humor*, *social sharing*, *social motivators*, *negative responses* and *self-disclosure*. Evidence about these signs, deriving from linguistic analysis of the user moves, is combined with information about the user's personality traits, the context and the previous move of the Agent. The model is employed to guess how the social attitude of the user evolves during a dialog by also considering the dialog history. The conversational agent's behavior is adapted accordingly as follows:

– at the beginning of interaction, the model is initialized by propagating evidence about stable user features;

– every time a new user move is entered, its linguistic features are analyzed and the resulting evidence is introduced and propagated in the network, together with evidence about the context;

– the new probabilities of the signs of social attitude are read and contribute to formulating the next agent move; the probability of the dynamic variable representing social attitude supports revising high-level planning of the agent's behavior, according to the goal of adopting a successful persuasion strategy. More details about this model and about its extension to acoustic signs of social attitude can be found in (Carofiglio et al., 2005) and (de Rosis et al., 2007).

Social-Mind has some similarities with the work described in Shaikh et al (this Volume): in that system, semantic parsing techniques are combined with a set of rules derived from the OCC classification of emotions (Ortony et al., 1988) to '*sense affective information*' from texts. The semantic parser produces a computational data model for every sentence in the form of triplets: this output is used for setting the value of the cognitive variables which act as antecedents of rules describing emotions triggering. The method allows sensing the affective content carried by every sentence which is provided as an input for the system.

In figure 3, *Emotional-Mind* and *Social-Mind* are compared with the models by Ball and Breese (2000), and Conati and McLaren (2005): the figure shows that the four models differ in the aspects of the 'emotion activation and expression' cycle they consider and in the grain size of knowledge represented. Some peculiarities make *Emotional-mind* and *Social-mind* two models of some value, especially in the context of affective dialogues. In particular: their ability to represent the role of personality and social relationship in emotion activation and time decay, the ability of *Emotional-Mind* to represent the various ways in which emotions may mix up with different intensities and, finally, its potential of being used, as well, for interpreting the reasons why an emotion is displayed (Carofiglio and de Rosis, 2005).

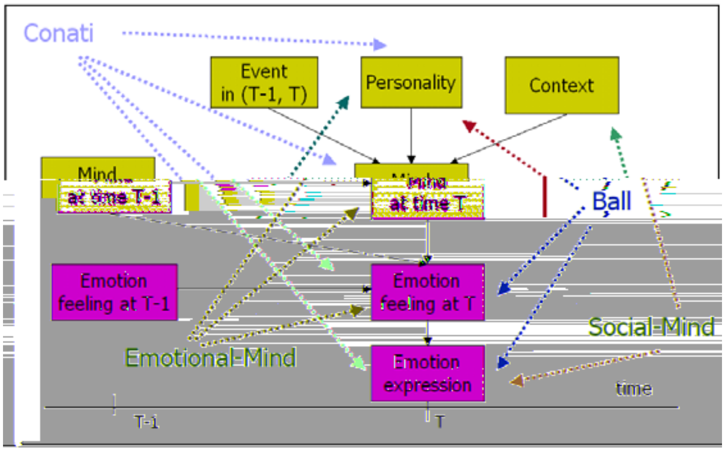

Fig 3: Comparison of some emotion models based on DBNs

6 Problems in Building Emotion Models as DBNs

The problem of how to estimate parameters when building probabilistic models is still a matter of discussion. In particular, in probabilistic emotion triggering models, the following questions are raised: What is the probability of an event? How much threatening for my life is this event? How important is, to me, to avoid this risk? and others. Similar considerations apply to defining variables involved in the recognition process. BN parameters can be estimated by learning them from a corpus of data (frequentist approach) or according to subjective experience or common sense (neo-Bayesian approach). To validate probabilistic emotion activation and expression models in Emotional-Mind, for example, a parameter sensitivity analysis (PSA) can be performed. This analysis investigates the effect of *uncertainty in the estimates of the network's parameters* on the probability of one or more nodes of interest, and discovers *critical parameters* which significantly affect the network's performance. Its purpose is to guide experts in making more precise assessments of the few parameters identified as 'critical', in tuning the network. Some notations:

– X and Z denote a node and its parent, x and z the states they (respectively) assume and p= P(X=x|Z=z) a conditional probability associated with the arc connecting them,

– h denotes a node or a subset subBN of nodes in the BN, and an evidence propagated in this node or in subBN,

– f(P(h|),p), f'(P(h,),p), g(P( ),p) are functions describing probability distributions when p is varied in the (0,1) interval.

PSA is based on the observation that the probability of ε is a linear function of any single parameter (conditional probability distribution) in the model; hence g(P( ),p) has the simple form y= p+ , where , are real numbers. This implies that the conditional probability distribution of a hypothesis given some evidence, when the considered parameter is varied, is the following:

f(P(h|),p) = f'(P(h,),p) / g(P( ),p) = p p

he coefficients of this linear function can be determined from the values it takes for two different value assignments to p. The two function values can be computed by propagating the evidence twice in the network (once for every value of p).

According to the method suggested by (Coupé and Van der Gaag, 2002), given a Bayesian network model with its parameters, a hypothesis (node of interest with its state) and a set of evidences, the task of sensitivity analysis is to determine how sensitive is the posterior probability of the hypothesis to variations in the value of some predefined parameters. This analysis can be simplified by considering changes in the value of a single parameter.

Hence, given one of the BNs of *Emotional-Mind*, if the model is used in a *prognostic mode*, the aim is to investigate which emotion is felt in a given situation. In this case:

– The hypothesis is the node which represents the felt emotion at time T (e.g. node (`Feel M HappyFor(C)`) in figure 1)

– The observed nodes are an appropriate combination of the emotional state at time T-1, intrinsic variables, influencing factors and context-related variables.

PSA may be applied, for example, to the activation of HappyFor in figure 1 (subnetwork during the time interval [T1,T2]), where:

– the hypothesis is `(Feel M HappyFor(C))` and the evidence set is ={(`Bel M Hap(GrowsUp-C)`), (`WarmRelationship M C)}`

– the parameter of interest is:
p1=P(`(Bel M Ach(GoodOfC))`|`(Bel M Hap(GrowsUp-C)),(Bel M (Goal C Hap(GrowsUp-C)),(Goal M Hap(GrowsUp-C))`).

That is, the goal is to estimate the posterior probability that M is feeling happy-for after observing that her lovely little child C is able to cycle by himself, depending on the initial assessment of the variables which are intrinsic to triggering the happy-for: M's belief and goal that her child is growing up, together with her belief that the mentioned goal is a valid one also for her child.

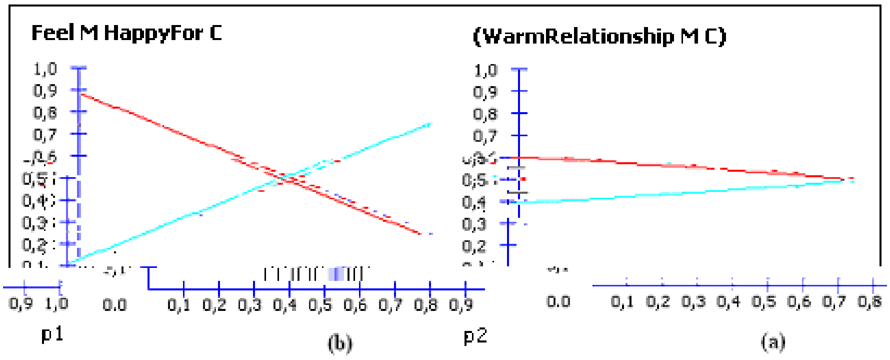

Fig.4. Parameter sensitivity analysis for the happy-for model: in a), the BN is used in a prognostic mode; in b), it is used in a diagnostic mode.

Figure 4a shows the sensitivity function f(P(h|),p) for this example[2]. This function is used to compute a belief value for the two states of the hypothesis variable: `(Feel M HappyFor(C))`= *yes/no*. The two lines represent how sensitive are the states of this variable to changes in the value of the parameter p1. The graph shows that the most likely state of the hypothesis variable changes when p1 = .6. It shows, as well, that the probability P(`(Feel M HappyFor(C))` |) is rather sensitive to variations in the value of p1, as one would expect; this reflects the fact that this parameter involves variables (with their states) which are intrinsic to triggering happy-for.

If the model is used in a *diagnostic mode*, to interpret the reason why an emotion was displayed, then:

- the hypothesis is one of the nodes who represent an intrinsic variable, an influencing factor or a context-related variable;
- the observed node is the felt emotion.

Figure 4b shows the sensitivity function f(P(h|),p) when the hypothesis is `(WarmRelationship M C)` and the initial evidence set is ={`(Feel M HappyFor(C))`}. This time, we are interested in the posterior probability of a (hypothesized) cause of the observed emotion of happy-for in M. In this case, p2=P(`(Bel M Goal C Hap(GrowsUp-C))`|`(Bel M Des(GrowsUp-`

[2] The graph is plotted with Hugin. www.hugin.com

`C)))` is the parameter of interest. The sensitivity function computes a probability value for the two states of the hypothesis variable, `(WarmRelationship M C)` = *yes/no*. The graph shows that the probability P(`(WarmRelationship M C)|`) is rather insensitive to variations in the value of p2.

In the discussed example, if the BN is used in a diagnostic mode the results of propagating evidence are rather insensitive to variations in the value of the considered parameter of the model. This is due to the structure of the network and also to its use (which observed nodes, which nodes of interest). If, on the contrary, the BN is used in a prognostic mode, sensitivity analysis helps in discovering parameters with a significant impact on the network's performance.

This analysis leaves room for further investigation about the quality of emotion models, by applying evidence sensitivity analysis (Coupé and Van der Gaag, 2002). This analysis is aimed at answering questions like: '*what are the minimum and the maximum probabilities produced in the hypothesis when observing a variable?*', '*which evidence acts in favor or against a given hypothesis?*', '*which evidence discriminates a hypothesis from an alternative one?*', and '*what-if a given variable had been observed with a different value?*'. Knowing the answers to these and similar questions may help to explain and understand the conclusions reached by the model as a result of probabilistic inference, and hence to evaluate the predictive value of emotion models.

7 An Open Problem: Cognitive Inconsistency

Irrespectively of the formalism adopted and of the grain size in knowledge representation, computational models of emotions are built on the hypothesis of *consistency* among the variables included in the model: variables that represent appraisal of the environment (occurred, occurring or future events), cognitive aspects that translate these variables into 'attitudes' (first and second-order beliefs, goals, values, etc) and emotions that are presumably activated. Various psychological studies are, however, seriously discussing this consistency hypothesis. In the following, two of them are considered, which give a slightly different interpretation of this phenomenon.

Goldie denotes with misleading emotions the emotions that are not useful in picking up saliences in the environment and enabling quick and effective action, as (on the contrary) emotions are presumed to do (Goldie, in press). To Brady, *recalcitrant emotions* are those emotions which involve a conflict between an emotional perception and an evaluative belief: the subject perceives his situation to be thus-and-so, while believing that it is not thus-and-so (Brady, 2007). For instance, a recalcitrant bout of fear is one where someone is afraid of something, despite believing that it poses little or no danger: he may believe both that flying is dangerous and that it is perfectly safe or (in the mother-child example introduced in Section 5), that running with a bike is safe while believing that it is not. Interpretation of the reasons of this contradiction differ in the two authors. Goldie follows the 'dual process' theory (Sloman, 1996), according to which our perception and response to emotional situations would be processed through two routes: (i) a *fast and frugal* route (also called 'intuitive thinking') which involves imagination, operates fast and uses limited resources and speed of processing, and (ii) a *more complex, slower route* (also called 'deliberative thinking'), whose function would be to operate as a check or balance on intuitive thinking. The dual process theory acknowledges the possibility that the two routes do not work in perfect agreement: therefore, it may happen that some emotions resulting from intuitive thinking *mislead us*, and that deliberative thinking does not succeed in correcting them. Goldie' hypothesis is that intuitive thinking would be performed through some *heuristics* that were built after environmental situations humans had to face in their past history. Changes in environmental conditions (that he calls *environmental mismatches*) would then be responsible for producing misleading emotions, which conflict with deliberative thinking, and that this complex and slower route is not able to detect and correct. Rather than accepting the hypothesis of recalcitrant emotions as 'irrational', Brady proposes a positive role for this contradiction, as a means to facilitate the processing of emotional stimuli: even if our deliberative thinking recognizes the felt emotional state as 'unreasonable', our emotional system would ensure that our attention remains fixed on the dangerous objects and events, thus checking them and facilitating a more *accurate* representation of the danger (or the insult, the loss, for emotions different from fear). In this interpretation, the excessive and recalcitrant emotion of fear in the mother-child example (section 5) is functional to her checking carefully that her child does not adopt a dangerous attitude in cycling.

This contradiction between emotional and cognitive state is one of the cases of *cognitive dissonance* that was originally described by Festinger (1957). To this author, 'cognitions' are element of knowledge, such as beliefs, attitudes, values and feelings (about oneself, others or the environment); dissonance may occur among any of these attitudes. His definition of dissonance is quite strong, as cognitions are said to be 'dissonant with one another when they are *inconsistent* with each other or *contradict* one another': therefore, a logical view of contradiction. In Brady and in Goldie, on the contrary, 'weak' contradictions may also occur, as they may involve (again, for instance, in the case of fear) the estimation of a 'degree of dangerousness' that influences a 'degree of inconsistency' or incongruence with other attitudes.

Festinger's theory, subsequently elaborated by Harmon Jones (2000), was focused on the study of the (negative) emotions that result from becoming aware of a state of cognitive dissonance, and on how these negative effects can be reduced. In the context of this Chapter, this theory is evoked in order to highlight that influential psychological theories exist, according to which the human mind cannot be assumed to be internally consistent. And this problem should not be ignored in building computational models of emotion activation, irrespectively of the formalism employed. The following *alternatives* then arise in building these models: a) *don't make room for the possibility of conflict:* but then an important part of emotional life is eliminated; b) *make room for the possibility of conflict:* but then the problem should be considered of how to emulate such a kind of representational state. The second alternative might be implemented by representing the dual process with two separate models, one for intuitive and one for reflective thinking (as envisaged in Cañamero, 2005). In this case, the cognitive component should represent the ability to correct errors introduced by fast and frugal intuitive algorithms; however it should leave space, at the same time, for occurrence of 'misleading' or 'recalcitrant' emotions and should deal with them. This is a quite demanding and still not tackled challenge.

Emotion models that deal with uncertainty, however, do enable representing cognitive-emotional contradictions, although in a quite simplified way. Several typical examples of weak contradiction between cognition and negative emotions may be found in the ISEAR Databank[3]. In the following one of these example, a case of fear, is taken as starting point for further reasoning:

> *"If I walk alone in the night, it might happen that I will be attacked. I feel fearful, even though I don't believe that it is likely that I will be attacked".*

In this example, the situation is hypothetical (or was possibly experienced in the past): fear is felt in spite of a *low* or *very low* danger ('it is not likely that I will be attacked', 'it might happen').

If:

A denotes the agent on which reasoning is performed

`WalkAloneInTheNight-A` is the considered event occurring to `A`

`Attacked(A)`, `BadConsequences(A)`, are formulas denoting the state of the world after the event occurred (A is attacked, bad consequences occur to A)

? denotes an 'uncertain implication',

the activation of fear may be represented (in probabilistic terms) as follows [4]:

---

[3] In the 1990s, a large group of psychologists collected data in a project directed by Klaus R. Scherer and Harald Wallbott (Geneva University). Student respondents, both psychologists and non-psychologists, were asked to report situations in which they had experienced all of 7 major emotions (joy, fear, anger, sadness, disgust, shame, and guilt). The final data set thus contained reports on these emotions by close to 3000 respondents in 37 countries on all 5 continents. http://www.unige.ch/fapse/emotion/databanks/isear.html

[4] In this model, A's attitudes that bring her to feeling fear are represented with a slightly different formalism than in figure 1.

`WalkAloneInTheNight-A ? Bel A Attacked(A)`, with:
P(`(Bel A Attacked(A))|WalkAloneInTheNight-A`) = p1, *probability that A believes she will be attacked when walking alone in the night;*

`Bel A (Attacked(A) ? BadConsequences(A))`, with:
P(`BadConsequences(A)|Attacked(A)`)=p2, *probability that A associates to bad consequences of being attacked;* degree of danger of these consequences = d;

`(Bel A BadConsequences(A)) ? (Bel A Thr(GoodOfSelf))` , with P(`(Bel A Thr(GoodOfSelf))|(Bel A BadConsequences(A))`) =p3, *probability that A perceives her goal of self-preservation as threatened, when bad consequences of being attacked occur;* w = weight attached by A to GoodOfSelf;

`(Bel A Thr(GoodOfSelf)) ? (Feel A Fear)`, with
P(`(Feel A Fear) | (Bel A Thr(GoodOfSelf)`)) p3, *probability that A feels fearful after perceiving her goal of self-preservation as threatened;* the intensity of fear will be a function of p3 but also of w.

In applying this model to different situations and different subjects, it may happen that the intensity of fear felt by A is not too low, although A considers unlikely to be attacked (when p1 is low but p2, p3, d and w are high): this will simulate a case of mild cognitive inconsistency. There will be, however, also cases in which, in the same situation (same value of p1), A will not feel fear, if p2, p3, d and w are low: for instance, if A is persuaded that attacks in the area in which she is walking are not very dangerous and if she is a brave person. In these cases, cognitive inconsistency will not occur.

Overall, weak contradictions may therefore be produced by the following factors:

a *overestimating conditional likelihoods* (p1, p2 and p3 in our example): as Kahneman et al (1982) pointed out, in subjectively estimating probabilities humans apply some 'quick-and-dirty' heuristics which are usually very effective, but may lead them to some bias in particular situations. '*How dark is the place in which I'm walking*', '*whether there is some unpleasant noise*', or '*whether I'm nervous*', are examples of factors that may bias this estimate;

b *overweighting of the losses due to the negative event* (d in the example above): the 'risk aversion' effect may bring the subject to overweight the losses and to be distressed by this perspective;

c *overweighting the goal of self-preservation* (w in the example above), due to personality factors.

These contradictions can be represented in emotion models by trying to emulate the way humans make inferences about unknown aspects of the environment. Uncertainty will be represented, in this case, with some algorithm that makes ‘near-optimal inferences’ with limited knowledge and in a fast way, like those proposed by Gigerenzer and Goldstein (1996). Alternatively, as in the representation considered in this Chapter, they can be built on probability theory. In this case, models will have to include consideration of context variables that might bring the subject to over or underestimate conditional likelihoods, and losses or gains due (respectively) to negative or positive events. That is, rather than assigning fixed parameters to the dynamic belief networks (as it happens in *Emotional-Mind*), conditional probability distributions should be driven by selected context variables. This is another reason for including context in emotion activation models, a direction followed in *Emotional Mind* and *Social Mind* (section 5).

Acknowledgments This work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotions (EC Contract 507422). The authors owe many of the ideas about cognitive-emotional contradiction, and their role in emotion modeling, to Professor Peter Goldie (University of Manchester), with which one of them had interesting and enlightening discussions in the scope of HUMAINE.

References

Andersen, P A and Guerrero, L K (1998) The brightside of relational communication: interpersonal warmth as a social emotion. In Andersen, P A and Guerrero, L K. (Eds.) Handbook of Communication and Emotions. Research, theory, applications and contexts. Academic Press

Arnold, M B (1960) Emotion and personality. Columbia University Press, New York (1960).

Ball, G and Breese, J (2000). Emotion and personality in a conversational agent. In S. Prevost, J. Cassell, J. Sullivan & E. Churchill (Eds) Embodied Conversational Agents. Cambridge, MA: The MIT Press, 89-219

Brady, M (2007) Recalcitrant emotions and visual illusions. American Philosophical Quarterly 44, 3: 273-284.

Cañamero, L (Coordinator) (2005) Emotion in Cognition and Action. Deliverable D7d of the HUMAINE NoE.

Carberry, S and de Rosis, F (2008) Introduction to the Special Issue of UMUAI on ‘Affective Modeling and Adaptation’ 18, 1-2, 2008, 1-9.

Carofiglio, V and de Rosis, F (2005) In favour of cognitive models of emotions. In Proceedings of the Workshop on ‘Mind Minding Agents’, in the context of AISB05.

Carofiglio, V de Rosis, F and Novielli, N (2005) Dynamic User Modeling in Health Promotion Dialogues. In Tao, J, Tan, T and Picard, R W (Eds) Affective Computing and Intelligent Interaction, First International Conference, ACII 2005. Springer, LNCS 3784: 723-730

Carofiglio, V, de Rosis, F and Grassano, R (in press). Dynamic models of mixed emotion activation. In Canamero, L and Aylett, R (Eds) Animating expressive characters for social interactions. John Benjamins Publ Co.

Conati, C (2002). Probabilistic assessment of user’s emotions in educational games. Applied Artificial Intelligence [Special Issue on ‘Merging cognition and affect in HCI’].16: 555– 575.

Conati, C and MacLaren, H. (2005) Data-driven refinement of a probabilistic model of user affect. In L. Ardissono, P.Brna and A.Mitrovic (Eds): User Modeling 2005. Springer LNAI 3538, 40-49.

Coupé, V M H and Van der Gaag, L C (2002) Properties of Sensitivity Analysis of Bayesian Belief Networks. Ann Math Artif Intell, 36, 4:323-356.

Cowie, R (2006) Emotional life, terminological and conceptual clarifications. Available at the HUMAINE Portal: http://emotion-research.net/deliverables/D3i final.pdf.

de Rosis, F, Pelachaud, C, Poggi, I, De Carolis, B and Carofiglio, V (2003) From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. In International Journal of Human-Computer Studies 59 (1/2) 81–118.

de Rosis, F, Batliner, A, Novielli, N and Steidl, S (2007) 'You are sooo cool Valentina!'. Recognizing social attitude in speech-based dialogues with an ECA. In Paiva, A, Prada, R and Picard, R W (Eds) Affective Computing and Intelligent Interaction, Second International Conference, ACII 2007. Springer, LNCS 4738: 179-190

Elliott, C. and Siegle, G. (1993). Variables influencing the intensity of simulated affective states. In Reasoning about Mental States – Formal Theories & Applications. Papes from the 1993 AAAI Spring Symposium [Technical Report SS-93-05]. Menlo Park, CA: AAAI Press 58-67

Ekman, P (1999) Basic Emotions. In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion. Sussex, U.K.: John Wiley & Sons, Ltd. 45-60

Festinger, L (1957) A theory of cognitive dissonance. Stanford University Press.

Frijda N H and Mesquita, B (2000) The influence of emotions on beliefs. In Frijda , N H Manstead, A S R and Bem, S (Eds) Emotions and beliefs: How feelings influence thoughts. Cambridge University Press.

Gigerenzer, G and Goldstein, D G (1996) Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review, 103, 4: 650-669.

Goldie, P (in press) Misleading emotions. In D G Brun, U Doguolu and D Kuenzle (Eds) Epistemology and Emotions, Ashgate Publishing

Gratch, J and Marsella, S (2001) Tears and Fears: Modelling emotions and emotional behaviors in synthetic agents. In Proceedings of the 5th International Conference on Autonomous Agents New York: ACM Press 278-285

Harmon Jones, E (2000) A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes. In Frijda, N H Manstead A S R and Bem, S (Eds) Emotions and Bliefs, How feelings influence thoughts. Cambridge University Press 185-211.

James, W (1984) What is an emotion? (1984) Mind 9: 188- 205

Kahneman, D Slovic, P and Tversky,A (1982) Judgment under uncertainty, heuristics and biases. Cambridge University Press

Izard, C E (1993) Four systems for emotion activation: cognitive and non cognitive processes. Psychological Review 100, 1:68-90

Lazarus R S (1991) Emotion and adaptation. New York: Oxford University Press.

Mc Crae, R and John, O P (1992) An introduction to the Five-Factor model and its applications. Journal of Personality 60:175-215.

Miceli, M, de Rosis, F and Poggi, I (2006) Emotional and non emotional persuasion. Applied Artificial Intelligence: an International Journal.

Nicholson, A E and Brady, J M (1994) Dynamic belief networks for discrete monitoring. IEEE Transactions on Systems, Men and Cybernetics, 24,11:1593-1610.

Oatley, K. and 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

Pearl, J (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.

Picard, RW (1997) Affective Computing. The MIT Press.

Plutchik, R (1980) A general psycho-evolutionary theory of emotion. In Plutchik, R and Kellerman, H (Eds) Emotion: theory, research, and experiences New York, Academic press. 1:3-33.

Poggi, I and Pelachaud, C (1998) Performative Faces. Speech Communication, 26: 5-21.

Poggi I (2005) The goals of persuasion. Pragmatics and Cognition. 13,2: 297-336.

Polhemus, L, Shih, L-F and Swan, K (2001) Virtual interactivity: the representation of social presence in an on line discussion. Annual Meeting of the American Educational Research Association

Prendinger, H, Descamps, S and Ishizuka, M (2002) Scripting affective communication with life-like characters. Applied Artificial Intelligence [Special Issue on ‘Merging cognition and affect in HCI’] 16 (7-8): 519-553.

Prendinger, H and Ishizuka, M (2005) The empathic companion. A character-based interface that addresses users' affective states. AAI, 19:267-285

Reilly, N (1996) Believable social and emotional agents. Ph.D. diss., School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.

Russell, J A (2003) Core affect and the psychological construction of emotion. Psychological Review 110:145-172.

Russell, S and Norvig, P (1995) Artificial Intelligence: A Modern Approach.. Morgan-Kaufman

Scherer, K R, Wranik, T, Sangsue, J, Tran, V, and Scherer, U (2004) Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction pattern. Soc. Sci. Info.: 43,4:499--570.

Shaikh, M A M, Prendinger, H and Ishizuka, M: A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text, this book

Sloman, SA (1996) The empirical case for two systems of reasoning. Psychological Bulletin, 119, 1:3-22

van Reekum, C M, and Scherer, K. R. (1997) Levels of processing in emotion-antecedent appraisal. In Matthews, G (Ed.), Cognitive science perspectives on personality and emotion Amsterdam: Elsevier Science 259-300