Draft Chapter of HUMAINE HANDBOOK, Springer, in press

Cognitive Evaluations And Intuitive Appraisals: Can Emotion Models Handle Them Both?

Fiorella de Rosis[1] *, Cristiano Castelfranchi*[2]*, Peter Goldie*[3] *and Valeria Carofiglio*[1]

[1] Department of Informatics, University of Bari
Via Orabona 4, 70124 Bari, Italy
{carofiglio,derosis}@di.uniba.it

[2] Institute of Cognitive Sciences and Technologies, National Research Council
via San Martino della Battaglia 44, 00185 - Roma Italy
cristiano.castelfranchi@istc.cnr.it

[3] School of Social Sciences, The University of Manchester
Bridgeford Street, Manchester, UK
peter.goldie@manchester.ac.uk

1. Introduction

Affective interaction is a sub-kind of social interaction, which entails an interaction between minds, and a reaction to the mind behind a perceived behavior. A given behavior can have completely different meanings, and we actually react to these meanings rather than directly to the perceived behavior. B is giving one dollar to A: is she 'paying' A? Or 'loaning' him one dollar, or 'giving him back' this money, or 'offering' it, or 'handing it out'? Or does she give A one dollar so that he can buy something for her? … and so on .

An advantage of the belief-desire-intention (BDI) paradigm -in a broad sense, not as a specific AI architecture- with emotions and affective interaction is the fact that an agent A should not necessarily respond to an emotion of another agent B as a global expressive input, and with an emotional reaction in its turn.

- First, A can *perceive, recognize* (read) something *behind* the emotion: the beliefs, desires, intentions and feelings of B, which motivate or constitute it (Carofiglio and de Rosis, 2005).
- Second, rather than with an emotional, 'mirrored' reaction, A can *react* to the emotion of B with a belief about her mental state, and perhaps without any specific emotion towards this. For example, rather than reacting with pity to B's

display of pain or anger, A may think "She clearly *believes* that John has taken her dog" or "She clearly has the *goal* to be approved/esteemed by them!".

- Third, A can *specifically* react to those beliefs, desires, intentions or feelings. For example, if B shows some pain A can react either empathically or not, by saying: "*Are you suffering a lot?*" (to show a sympathetic attitude) or "*You are wrong; it is not as you believe*" (to try to stop an unmotivated pain) or "*It is not fair; you should be angry, not sorrow!*" (to attempt to change B's attitude from that of a victim to an active, angry, indignant one) or "*You shouldn't care of these stupid things*" (to reduce B's pain by instilling a bit of indifference, superiority in her), etc. In those cases, what triggers in A an emotion towards B or his own emotion are precisely A's beliefs about B's attitudes: that is, about beliefs, desires, intentions and feelings he ascribes to B.

In interacting with other (human or artificial) agents, we have a causal theory and a mental model of their emotion (including their psychological processes) and we recognize and interpret this emotion not in terms of what we see, but in terms of what we cannot see: their mind.

2. Cognitive Evaluations vs Intuitive Appraisals

In the psychological literature on emotions, there is a systematic confusion between two kinds of "evaluation":

- A *declarative or explicit* form of evaluation, that contains a judgment of a means-end link, frequently supported by some reason for this judgment, relative to some "quality" or standard satisfaction. This reason-based evaluation can be discussed, explained, based on arguments, as well as the goal of having/using the well-evaluated entity (which is the declarative equivalent of "attraction"). See the "motto" (of Aristotelian spirit): *"it is pleasant/we like it, because it is good/beautiful"*A *non-rational* (but adaptive) evaluation, not based on justifiable reasons; this is a mere "appraisal", based on associative learning and memory.

Castelfranchi (2000) proposed to distinguish between *appraisal* -the unconscious or automatic, implicit, intuitive orientation towards what is good and what is bad for the organism- and *evaluation* -the cognitive judgments relative to what is good or bad for someone (and why). Appraisal is therefore an associated, conditioned somatic response; it has a central component and involves pleasure/displeasure, attraction/repulsion; it is a preliminary, central and preparatory part of motor response. This response can be merely central or more complete, involving overt motor or muscle responses or somatic emotional reactions. It is automatic, and frequently unconscious**;** it is a way of 'feeling' something, thanks to its somatic (although central) nature; it gives valence to the stimulus by making it attractive or repulsive, good or bad, pleasant or disagreeable; it has intensionality, as the association/activation makes nice or bad, fearful or attractive what we feel about the stimulus. When it is a response just to the stimulus, appraisal is very fast, primary: it anticipates high level processing of the stimulus -like meaning retrieval- and even its recognition -which can be subliminal-. However, associative, conditioned,

automatic responses to high level representations may occur as well: to beliefs, to hypothetical scenarios and decisions (Damasio, 1995), to mental images, to goals, etc.

We propose to change our usual view of cognitive 'layers', where association and conditioning are only relative to stimuli and behaviours, not to cognitive mental representations.While not all emotions suppose or imply a cognitive evaluation of the circumstances, any emotion as a response implies an appraisal in the above mentioned sense: it implies the elicitation of a central affective response involving pleasure/displeasure, attraction/repulsion, and central somatic markers if not peripheral reactions and sensations. This is what gives emotions their 'felt' character.

Evaluation and appraisal can also derive from each other. Merely affective reactions towards some event can be verbalised, translated into declarative appreciations. The opposite path - from a cold evaluation to a hot appraisal - is also possible, especially for personal, active, important goals, and in particular for felt kinds of goals like needs, desires, etc. The appraisal of an event produces a feedback on the beliefs and confirms them: *"Since I'm afraid, it should be dangerous. I was right!"*

Evaluations do not necessarily imply emotions: not any belief about the goodness or badness of something necessarily implies or induces an emotion or an attraction/rejection with regard to that something. Cold evaluations also exist. Evaluations are likely to have emotional consequences if they are about our own goals, if these goals are currently active and are important.motions do not necessarily imply evaluations: attraction or rejection might be considered *per se* as forms of evaluation of the attractive or repulsive object: we view attraction and rejection as pre-cognitive implicit evaluations that we call "appraisal".

To summarise: mind should be incorporated into emotion models with a structured and detailed representation of beliefs (evaluations, expectations,...) and goals (intentions, needs, desires), by going deep into the understanding of their specific relationships with 'emotional feelings', that is with the body appraisal and response.

2.1 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).

Another, less relevant relation between emotions and goals is that - given that emotions are pleasant or unpleasant experiences - to feel or not to feel an emotional state (for example, excitement, funny; or fear, boredom, ..) can *per se'* become a goal to the individual. Not necessarily we always avoid negative emotions (like fear or disgust); there are movies specialized for these experiences. There are even individuals or situations where the goal is 'not to feel emotions at all', 'not to be moved or disturbed'; or whose goal is to feel some emotion, to be aroused, just in order to 'feel alive'.

2.2. 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 we 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, in their turn, 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 (we will come back again on this issue in Section 6). 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.

In complex emotions, like guilt, jealousy or envy, a specific set of beliefs not only is necessary for triggering the emotion as interpretation, evaluation of the event or as its prediction. Beliefs characterize, in this case, the mental attitude of that emotion; they persist during the emotion and, if they are invalidated, the emotion is stopped, reduced or changed. They are explicit, that is non necessarily conscious or under attention: they are subject to possible 'argumentation' for example from a friend of us, or of manipulation from ourselves, in order to reduce the emotion or change it.

To defend oneself from pity or guilt one can, for example, introduce a belief that the harm and the possible suffering is 'fair', is due; for example: "*The guy was deserving that punishment, it is his fault*". Or - to transform envy - one may think that the other's enviable condition "is unfair, is an unmotivated privilege", that "it is a case of injustice"; this belief will transform the despicable, hidden and passive 'envy' into a noble and open 'sense of injustice', 'indignation' (Miceli & Castelfranchi, 2007).

To some authors (Feldman-Barrett, et al. 2005; Miceli et al. 1997), for typical and complex human emotions even a belief of 'recognition' or 'categorization' of that mental and feeling state as such an emotion is necessary for specifically and fully feeling that emotion: this is culturally specific, and we have learnt to discriminate it from similar emotions also thanks to their cultural label.

3. How Can We Apply These Theories To Build Emotion Models?

In trying to formalize a computationally tractable model of emotion, we 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 likely 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 interaction.

3.1. Aspects to include in a computational emotion model

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 (1) the link between *situation* characteristics (which event -and consequences- which social aspects -role of individuals involved in the situation- which individual's personality traits) and emotion components (appraisal variables and other internal dispositions -beliefs and goals), and (2) the mutual links between emotion components themselves (van Reekum & Scherer, 1997).

b. Emotional dispositions

A second major source of difference is in individual emotional dispositions. An emotional disposition is a disposition to have certain kinds of occurrent thoughts and feelings towards a certain kind of thing—that is, towards a *focus* (Helm 2001; Goldie forthcoming). So John's fear of cats is an emotional disposition, with cats as the focus. The presence of this disposition will explain why John feels fear on seeing a cat, or on being told that the house he is going to for dinner is full of cats; whereas other people will not feel fear in these situations. This isn't a personality trait (it's too focused for that), not is it obviously a 'different structure of beliefs and goals' – at least if it is the latter, then only derivatively so, in the sense that John will have a goal of avoiding cats – but only because he is afraid of them (in the dispositional sense).

c. 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 et al., 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 a lower threshold in emotion feeling (Ortony et al., 1988): 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, etc. 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.

d. 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* (resources needed to make the prospected event happen or to prevent its happening) and the *realization* (degree to which the prospected event actually happens). The intensity of fortune-of-others emotions is affected by: presumed *desirability* for the other, *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 deserved what happened to him). Some *global intensity variables* also affect the intensity of emotions: *sense of reality* (how much the emotion-inducing situation is real); *proximity* (how much the individual feels psychologically close to the emotion-inducing situation); *unexpectedness* (how much the individual is surprised by the situation); *arousal* (how much excited the individual was before the stimulus).

e. Emotion mixing and oscillation: partially overlapping emotions

Emotions may co-occur or swing from each other: his can be due to overlapping in their cognitive backgrounds. In a sense, for example, *guilt* feeling towards a victim B 'contains' *pity*, as they have the same constituents: i) the belief that B received a serious harm/loss; ii) the idea that (because of this) he is suffering, will or might suffer; iii) an empathic disposition and feeling based on this perceived or imagined suffering; iv) the idea that such a harm and suffering is not fair, is not a right punishment for B and his conduct. This cognitive pattern elicits a goal, an impulse to care about B, to worry of his pain and condition, to succour and help him if possible. However, pity is the empathic and charitable feeling of the bystander, who is not responsible of such an unfair suffering. On the contrary, guilt is the affective attitude of those who feel responsible of that harm: so, the helping impulse not only reduces pity but also alleviates the sense of irresponsibility and guilt. Within the broader frame of guilt, the same generous impulse, based on the same beliefs, changes its 'flavor': in a sense, therefore, guilt towards a victim contains 'pity'; but in another sense - in the scope of a larger gestalt - it is no longer just pity.

Other examples of partial overlapping are *guilt* and *shame*, which overlap in important constituents though being focused on different goals. While what matters in shame is the others' opinion about us, our social 'image', what matters in guilt is the violation of a moral standard. In both cases, however, there are negative evaluations about B (since to be a morally bad guy is a negative judgment) as well as self-evaluations. This is why shame and guilt can coexist, or why we can oscillate from each other if we care about social esteem and the others' possible judgment. This does not mean that, when there is shame, there can be guilt: shame is not necessarily a 'moral' emotion: it can be an 'esthetic' one (Sabini & Silver, 1982; Castelfranchi and

Poggi, 1990), for example concerning our physical aspect, and not implying any responsibility at all. But in all cases we do not correspond to some evaluation standard, and we get negative judgments. In shame I have some inferiority, I lack something, I am a defective guy; in guilt I did something bad, I am a bad guy. The two emotions can be transformed into each other also because my impotence: my lacking something can harm or make suffer somebody; or – vice versa – my bad behavior can be due to some defect or to my being in an inferior position.

In considering the problem of emotion mixing, Picard (1997) proposed two metaphors: 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: 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, 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. 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* and emotions that *switch* from each other over time. 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. This idea of generative mechanism is very close to Castelfranchi's idea of overlapping cognitive background.

f. Role of uncertainty

In some emotion modeling systems, emotion intensity is measured in terms of a mathematical function combining the previously mentioned variables (Prendinger et al, 2002; Elliott et al. 1993) or of logic rules (Turrini et al, 2007). An alternative 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 generative mechanism of emotions, the intensity with which they are triggered, the time decay of this intensity and the transition through various emotional states. Due to the presence of various sources of uncertainty, *dynamic belief networks* (DBNs) are a good formalism to achieve this aim (Nicholson and Brady, 1994): in the next section we will focus our attention on some remarkable examples of this kind.

3.2. Main experiences

In a first attempt in this domain (Ball and Breese, 2000), a simple model is proposed in which emotional state and personality traits were characterized by discrete values along a small number of dimensions (valence & arousal, dominance & friendliness). These internal states are treated as unobservable variables in a Bayesian network model. Model dependencies are 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) simulates the emotion decay over time for a

specific set of emotions, according to the goal that generated them. A specific intensity threshold is defined, for each emotion to be triggered. Affective Reasoner (Elliott & Siegle, 1993) is a typical multi-agent model. Each agent uses a representation of both itself and the interlocutor's mind, evaluates events according to the OCC theory and simulates a social interactive behavior by using its knowledge to infer the other's mental state. Emile (Gratch & Marsella, 2001) extends Affective Reasoner by defining emotion triggering in terms of plans representation: the intensity of emotions is strictly correlated to the probability of a plan to be executed, which is responsible for 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) belong to the OCC classification; some personality traits are assumed to affect the student's goals. The grain size of representation is not very fine, as among the various attitudes that may influence emotion activation (first and second order beliefs and goals, values etc) only goals are 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 a probabilistic decision network: these networks include representation of events, agent's choices and utility functions. In Emotional-Mind, Carofiglio et al (in press) advocate for 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, to activate one or more *individual emotions* in the OCC classification. We will demonstrate the representation power of this modeling method in Section 4.2 by considering, in particular, the emotion of fear.

4. The case of expectation-based emotions

We will now focus our discussion on the category of 'expectation-based emotions'. In our view, '*Expectation*' is not a mere forecast (a belief about the future) (Castelfranchi & Lorini, 2003): it is a complex and unitary mental object including tree components:

- A *forecast* (more or less certain, strong) about a future state or event Ev by agent A: (Bel A ◊Ev) [1];
- An '*expecting*' activity or disposition: A has the *goal of coming to know* whether his prediction was correct, whether the event actually happens:

  Goal A (Know-whether A Ev).

This is the basic meaning of 'expecting' and of the derived term 'expectation'. This is actually a specific kind of 'epistemic goal' and activity: an activity aimed at acquiring knowledge, the goal to 'know whether'.

[1] where ◊ denotes 'eventually', 'in a more or less near -and possible- future'

- But, why has A the goal to know whether her prediction is correct or not? Because the event is 'relevant' for A, because A is concerned by Ev. That is, a goal of A is ‘into question’ because of the -possible- event. Thus, the third ingredient of a full expectation is A's goal about Ev: (Goal A ¬Ev).

Overall: (Bel A ◊Ev) ∧ (Goal A (Know-whether A Ev)) ∧ (Goal A ¬Ev) [3]

This formalisation explains why currently *computers can make forecasts but cannot have expectations*; they are not 'concerned' and expecting for. A 'forecast' or 'prediction' is the result of a previous process of 'predicting', 'forecasting'; while 'expecting' is an activity or disposition which follows the formulation of the forecast and makes it an 'expectation'. To model the activity of monitoring, of knowledge acquisition that is typical of anticipation, computers should be modeled so as to become ‘involved’ about event evolution: they should be endowed with the ability, the interest, the curiosity to know how things will evolve. We willcome back to this point in Section 6.

There are various kinds of expectations (positive, negative, neutral and ambivalent (which can be defined relatively to A's goals:

- *positive expectation* (goal conformable): (Bel A ◊Ev) ∧ (Goal A Ev)
  A expects Ev to be eventually true, and wants it
- *negative expectation* (goal opposite): (Bel A ◊Ev) ∧ (Goal A ¬Ev)
  A expects Ev to be eventually true, and does not want it
- *neutral expectation:* (Bel A ◊Ev) ∧ ¬(Goal A Ev) ∧ ¬(Goal A ¬Ev)
  X expects Ev to be eventually true, and is explicitly indifferent about this
- *ambivalent expectation:* (Bel A ◊Ev) ∧ (Goal A Ev) ∧ (Goal A ¬Ev)
  X expects Ev to be eventually true, and on one hand he wants it while on the other hand he doesn’t. For instance, he may desire something and fear it at the same time.

4.1. The example of fear

The general function of fear is to avoid a (represented or not represented) harm, threatening event. There are, in our view, different kinds of fear: not all of them are 'anticipatory' emotions in strict sense.

(i) A first form of fear (the reaction of start and fright) can be due to a mere, very fast stimulus, before any imagination or realization of possible dangers, even before the full recognition of the triggering stimulus itself. This form of fear (that we might call '*stimulus-based*' or '*merely reactive*') is anticipatory only functionally; that is, the function of the induced reaction is oriented to the future, to a possible impending danger. However, it is not cognitively anticipatory, i.e. based on a prediction, a belief (an image) about the future event. 'Anticipation' means that our behavior is *coordinated with a future event* (cit); but this event is not necessarily explicitly represented in our mind or in the control system governing our behavior. "Prediction" is the explicit representation of a future event. Only some forms of anticipatory behaviors are based on prediction, on the explicit mental representation of a future event.

(ii) The most typical form of fear is *prediction-based*: it is activated by a prediction contrary to our goals, a threatening prediction or a perceived dangerous event Ev. A predicts/expects - on the basis of her reasoning (inferences) or of mere associative learning and activation - a future bad event. 'Bad' means contrary to her goals. This can be explicitly appraised, by A's judgment, with an explicit representation of the goal, or can be implicitly appraised as 'bad' (contrary to some goal) by the associated affective reaction or evocation of a 'somatic marker'.
Notice that in this case A does not perceive P from the environment: the danger is not there, it is imagined or inferred, it is coming. She is reacting to her mind, to a self-provided stimulus.

(iii) A third form of fear is in a sense not anticipatory, not relative to the future but to the past! This is a post-hoc fear about something which didn't actually happen, and that A does not expect to happen. This is a *counterfactual fear* due to the mere idea (inference, imagination) of what *might have happened* (but did not happen in fact). This fear (like the anticipatory one) gives rise to 'relief' for the escaped harm: in this case, however, that harm was not expected before. Given the unexpected event Ev, I realize (imagine) that something very bad might have happened, and I fill fear in front of such an idea (image), although I know that it did not happen, and I do not expect that it will happen. Also in this case it is the mere endogenously produced mental representation of something, which is not there, which terrifies me; but it is not a predicted event.

The first two forms of fear share their 'anticipatory' character, while the second and the third one share their 'imagination-based' ('allucinatory') character. Of course, evolutionary speaking, also the third form of fear has a function, and this is oriented to the future: but not as a short-term prediction and preparation/reaction. Its main function is learning: that is, remembering that those situations can be dangerous. Although nothing actually happened, I know that something might have happened; and, by learning from my imagination like from my actual experience, I will evoke my mental bad experience (the merely imagined harm) like a real harm, and will become prepared and cautious in those circumstances.

The three forms of fear may coexist: reactive fear can be due to a ‘false alarm’, which triggers counterfactual thinking or expectation of incoming dangers; or, the idea of what might have happened can make us cautious and fearful about what might have happened. In some cases, they may even contradict each other: a merely reactive fear may be felt while events contrary to own goals are not (yet) predicted: something in the periphery of our field of vision moves suddenly, we jump back or get goose-flesh much before or without forecasting any future danger.

4.2. A cognitive model of fear

To formalize how fear may be activated, we consider its prototypical form, the *prediction-based* one. Its cognitive skeleton is a *negative expectation*:

A worries about Ev, she doesn't like Ev, desires that Ev will not happen, would like to avoid Ev; however, she predicts and expects that Ev will happen.

$$(\text{Bel A} \Diamond \text{Ev}) \wedge (\text{Goal A} \neg \text{Ev}) \qquad [1]$$

or

A worries about ¬ Ev, she likes Ev, desires that Ev will happen, would like Ev; however, she predicts and expects that unfortunately Ev will not happen

$$(\text{Bel A } \neg\Diamond \text{Ev}) \wedge (\text{Goal A Ev}) \quad [2]$$

Fear occurs, as well, in ambivalent expectations: that is, when an event is taken into account as a possible outcome, it is desired and feared at the same time and one can focus on one side or the other, or anxiously oscillate from each other.

4.3. Fear in Emotional-Mind

We now describe how *cognitive models of prediction-based emotion activation* are built in *Emotional Mind* (de Rosis et al, 2003; Carofiglio et al, in press), and how such an emotion may mix up with other kinds of emotion.

In this model, emotions are activated by the belief that a particular important goal may be achieved or threatened: therefore, our simulation is focused on the change in the belief about the achievement (or threating) of the goals of 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 hthl, orpossibTi2(sed ostan(Ti2(por)]nte)2(079 Tc 0.3451 Tw 0 4591

achieving that goal which is, in its turn, a function of the agent's personality (in the example, how much the mother is attached to her child). The figure shows that, at the same time, fear may be activated in M by the belief that C might fall down by cycling because of his low experience: (Bel M ◊AtRisk-C) and (Bel M UnDesirable(AtRisk-C)). This may threaten the goal of self-preservation (Bel M ◊Thr-GoodOf-Self)

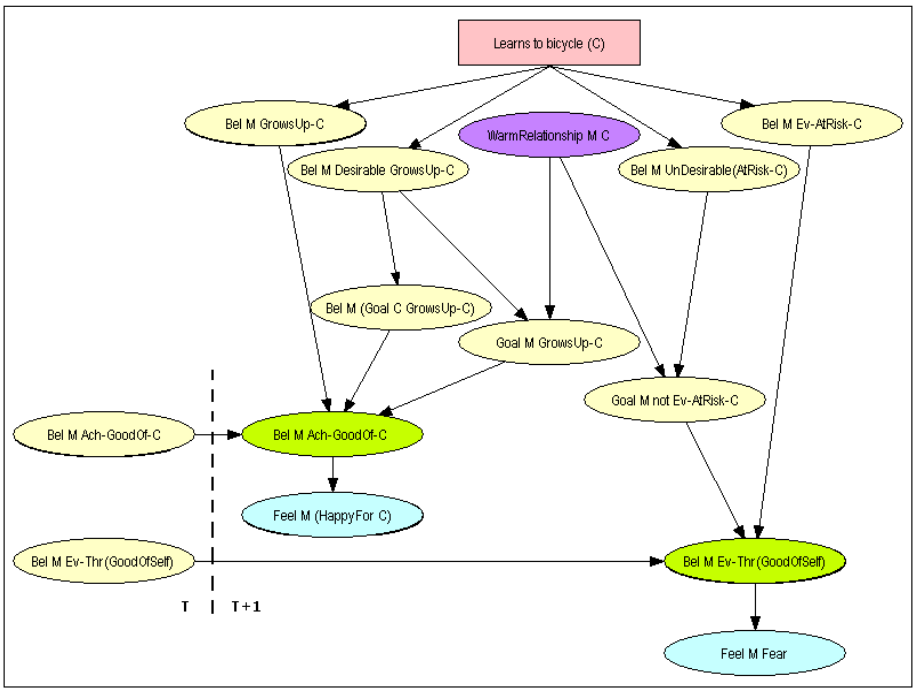

Fig.1. *An example of model for activation of fear and happy-for according to the 'tub-of-water' metaphor.*

5. 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, we we advocate for the need to investigate two different issues:

1. *Robustness of the model*: How sensitive are the results of a belief update (evidence propagation) to variations in the values of the parameters in the model (parameter sensitivity analysis)?
2. *Predictive value of the model*: Does the dynamic behaviour of the model reflect what an external, independent (and competent) source would expect? Evidence sensitivity analysis may help. This may, for instance, give answers to questions like what are the minimum and maximum beliefs produced by observing a variable, which evidence acts in favour or against a hypothesis, which evidence discriminates a hypothesis from an alternative one, and what-if an observed variable had been observed with a value different from the actual one. 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.

5.1. Parameter sensitivity analysis (PSA)

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. We consider, in particular, the case of discrete random variables and of changes in a single parameter value (one-way parameter sensitivity analysis). Some notations:

- X and Z are a node and its parent, and x and z are states they (respectively) assume and p= P(X=x|Z=z) is a conditional probability associated with the arc connecting them,
- h denotes a node or a subset subBN of nodes in the BN, and ε an evidence propagated in this node or in subBN,
- f(P(h|ε),p), f'(P(h,ε),p), g(P(ε),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(ε),p) has the simple form $y=\alpha p+\beta$, where α, β 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:

$$f(P(h|\varepsilon),p) = f'(P(h,\varepsilon),p) / g(P(\varepsilon),p) = \gamma p + \delta / \alpha p + \beta$$

The 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 (Coupe et al., 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. As we said, we simplify this analysis by considering changes in the value of a single parameter. If the BN is used in a *prognostic mode*, the model's aim is to investigate which emotion is felt in a given situation. In this case:

- The hypothesis is the node who 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.

5.2. Evidence sensitivity analysis (SE)

Evidence sensitivity analysis investigates how sensitive is the result of a belief update (propagation of evidence) to variations in the set of evidences (observations, likelihoods, etc.). Following (Jensen, 2001), if in the BN h and k denote hypothesis variables with state space (h_1, h_2, ..., h_n) and (k_1, k_2, ..., k_n), and $\varepsilon = \{\varepsilon_1, \varepsilon_2, ..., \varepsilon_m\}$ a set of evidences (findings) propagated in the BN, the impact of different subsets of ε on each state of the hypothesis variable can be studied, in order to determine:

- subsets of evidences $\varepsilon' \subseteq \varepsilon$ acting in favour or against the hypothesis;
- subsets of evidences $\varepsilon'' \subseteq \varepsilon$ discriminating between alternative hypotheses h and k.

In the case of cognitive emotion activation models, the aim is to investigate which emotion is felt in a given situation; then:

- the hypothesis h is the node which represents the emotion at time T; for instance, the node *Feel M Fear* in figure 1;
- the subsets of evidences are an appropriate combination of the emotional state at time T-1, intrinsic variables, influencing factors and context-related variables.

The impact of different subsets of evidences ε' on each state of the hypothesis variable can be investigated, to determine which combination of intrinsic and context-related variables and influencing factors acts in favour or against the hypothesis h: that is, which combination of parameters causes the triggering of a given emotion. This requires performing a propagation for every subset of the selected evidence ε'.

Another aspect that may be studied is the discrimination between two different (alternative) hypotheses. These may correspond, for instance, to emotions that cannot be activated at the same time, due to their non-compatible cognitive generation mechanisms: that is, due to the fact that some of the intrinsic variables, influencing factors and context-related variables in their activation subnets, take non-compatible values. Fear and Hope are examples of alternative emotions, because they are triggered by a different value of desirability of the occurring event and, therefore, by incompatible values of a node (intrinsic variable) they share in their generation subnets. Thus, if h is the hypothesis of interest (e.g., activation of fear) and k the alternative hypothesis (activation of hope), the discrimination requires to compare, in terms of bayesian likelihood ratio, the impacts of all subsets ε' of a set of evidence ε.

6. An open problem: Cognitive inconsistency

Let us introduce this topic with a premise about the difference between positive and negative emotions. While positive emotions are usually enjoyed as long as possible, and one tries to rehearsal them and persist, with negative emotions one is induced, by the malaise or pain, at going out 'here and now' from that mental state or at least at making it less severe, not just at avoiding the malaise and its eliciting situation in the future. So, in general a negative emotion activates a behavioral or mental activity of coping, some distraction or defense and reduction, some way out. What is remarkable is that frequently, if not normally, the activated impulses tend to alleviate or 'give vent' to the painful psychological state: so, escaping will reduce the fear; relieving the other will alleviate the felt pity or guilt for a victim, confessing will stop guilt feeling, aggression will give vent to the rage. The function of the impulse is not to reduce the subjective malaise: it is a specific function of that emotion, which is explained by the goal this emotion monitors. The function of pain-reduction is to learn and improve reaction in similar circumstances and to strengthen the induction of that behavior. This is one of the possible reasons on cognitive-emotional inconsistency, that we consider in this final Section.

6.1. Consistency in computational models of emotions

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. We will refer to two of them, 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 would be 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 we introduced in Section ~~5~~ 4), 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).

The excessive and recalcitrant emotion of fear in the mother of our example is functional to her checking carefully that her child does not adopt a dangerous attitude in cycling. This is very close to the kind of *counterfactual fear* that we considered in Section 4, due to the mere idea (inference, imagination) of what *might have happened* (but did not happen in fact). Its function is oriented to the future, not only as a short-term prediction and preparation/reaction but also to remember that those situations can be dangerous. Nothing actually happened but M, by learning from her imagination, evokes the merely imagined harm like a real harm, and becomes prepared and cautious.

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.

6.2. Some examples

We found several typical examples of weak contradiction between cognition and negative emotions in the ISEAR Databank[2]. To elaborate on the examples we

[2] 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

considered in the previous sections, we will cite and reason, here, on two cases of fear.

Ex 1:*"I was coming home from a relative's place and it was about 9.30/10 P.M. I felt slightly apprehensive when I got off the bus and started walking towards my place. I was confident that nothing would happen to me, yet there was this slight feeling of fear."*

Ex2: *"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"*

Ex 1 describes a real event occurred: a 'slight' fear was felt, in contradiction with a belief that the situation was *not* dangerous ('nothing would happen to me'). In Ex 2, the situation is hypothetical (or was possibly experienced in the past): this time, fear is felt in front of a *low* danger (it is not likely that I will be attacked).

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. We evoke this theory in the context of this Chapter 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. In our view, 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, to our knowledge, still not tackled challenge.

Emotion models that deal with uncertainty, however, do enable representing cognitive-emotional contradictions, although in a quite simplified way. Let us consider again the Example 2: *"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"*. If we denote: with A the agent on which we are reasoning, with a the action A is performing, with Ev an event, with t_h, t_k two time instants and with $\Rightarrow?$ an 'uncertain implication', the activation of fear may be represented as follows:

$$\{\text{Bel A } [\text{Does}(A,a,t_h) \Rightarrow ? \Diamond \text{Happens}(Ev,A,t_k)] \text{ (with } t_k > t_h) \text{ and Unpleasant}(Ev)\} \Rightarrow ? \text{Fearful } (A, t_h)$$

"If A believes that performing the action a at time t_h might produce the event Ev at a subsequent time t_k, and Ev is 'unpleasant', then A may feel fear".

Here, the likelihood of the unpleasant event Ev we are considering ('to be attacked') is, in fact, related to the conditional likelihood that this event will occur in a given situation ('walking alone in the night'); the same is true for the likelihood that dangerous consequences will occur, because of being attacked. If the first likelihood (of being attacked by walking alone in the night) is low, the likelihood of dangerous consequences will be low as well. In principle then, I should not feel fear; however,

the intensity of this feeling depends on how much dangerous are the consequences of being attacked and how much importance I give to my self-preservation. If I know that I'm risking my life or I tend to adopt a wise attitude (because of my personality or my past experiences), even a low likelihood will make me feel fearful. Hence, my contradictory state. Contradiction may be further increased by at least two factors:

a overestimating conditional likelihoods: 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: the 'risk aversion' effect may bring the subject to overweight the losses and to be distressed by this perspective.

This kind of bias can be represented by trying to emulate the way humans make inferences about unknown aspects of the environment. Uncertainty in computational models of emotion activation will be represented, in this case, with some algorithm capable of making 'near-optimal inferences' with limited knowledge and in a fast way, like those proposed by Gigerenzer and Goldstein (1996). Alternatively, as in the bayesian network representation we 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. Another reason for including context in emotion activation models, a direction we followed in our dynamic, uncertain models.

Acknowledgements

This work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotions (EC Contract 507422).

References

1. Arnold, M.B. (1960): Emotion and personality. Columbia University Press, New York (1960).
2. Ball, G. & Breese, J. (2000). Emotion and personality in a conversational agent. In S. Prevost, J. Cassell, J. Sullivan & E. Churchill (Eds), Embodied Conversational Agents (pp. 89-219). Cambridge, MA: The MIT Press.
3. M. Brady: Recalcitrant emotions and visual illusions. *American Philosophical Quarterly*, 44, 3, 2007, 273-284.
4. Cañamero,L. (Coordinator) (2005): Emotion in Cognition and Action. Deliverable D7d of the HUMAINE NoE.
5. Carofiglio,V. and de Rosis,F (2005): In favour of cognitive models of emotions. Proceedings of the Workshop on '*Mind Minding Agents'*, in the context of AISB05.
6. Carofiglio,V., F. de Rosis, and R. Grassano: (in press). 'Dynamic models of mixed emotion activation'. In: L. Canamero and R.Aylett (eds): *Animating expressive characters for social interactions*. John Benjamins Publ Co.

7. Castelfranchi, C. and Poggi, I. Blushing as a discourse: was Darwin wrong? In R. Crozier (ed.) *Shyness and Embarassement: Perspective from Social Psychology*, Cambridge University Press, N. Y, 1990.
8. Castelfranchi, C. (2000). Affective Appraisal vs Cognitive Evaluation in Social Emotions and Interactions. In A. Paiva (ed.) *Affective Interactions. Towards a New Generation of Computer Interfaces*. Heidelbergh, Springer, LNAI 1814, 76-106.
9. Castelfranchi, C., Lorini, E. (2003). Cognitive Anatomy and Functions of Expectations. In *Proceedings of IJCAI'03 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, Acapulco, Mexico, August 9-11, 2003.
10. 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.
11. 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, pp 40-49.
12. 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.
13. Damasio A. (1995) Descartes' Error. Emotion, Reason, and the Human Brain, Penguin Books - E Rutherford, NJ
14. de Rosis, F. Pelachaud, C., Poggi. I. De Carolis, N. & Carofiglio, V. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Intl. Journal of Human-Computer Studies, 59 (1/2), 81–118.
15. Elliott, C. & 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 (pp. 58-67) [Technical Report SS-93-05]. Menlo Park, CA: AAAI Press.
16. Feldman-Barrett, L. Niedenthal P.M., Winkielman, P., (Eds.) (2005) *Emotion And Consciousness, Guilford Pubn,. Boston*
17. Festinger,L.(1957) : *A theory of cognitive dissonance*. Stanford University Press.
18. Frijda N.H. and Mesquita,B. (2000): The influence of emotions on beliefs. In N.H.Frijda A.S.R.Manstead and S Bem (Eds). *Emotions and beliefs: How feelings influence thoughts.* Cambridge University Press, 2000.
19. Gigerenzer, G. and Goldstein, D.G. (1996): Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 4, 1996, 650-669.
20. Goldie,P.(in press): Misleading emotions. In *Epistemology and Emotions*, D G Brun, U Doguolu and D Kuenzle (Eds). Ashgate Publishing.
21. Goldie, P. Thick concepts and emotion. In *Reading Bernard Williams*m, D, Callcut (ed.), London: Routledge, forthcoming.
22. Gratch, J. & Marsella, S. (2001). Tears and Fears: Modelling emotions and emotional behaviors in synthetic agents. In Proceedings of the 5th International Conference on Autonomous Agents (pp. 278-285). New York: ACM Press.
23. Harmon Jones,E.(2000): A cognitive dissonance theory perspective on the role of emotion in the maintenance and change of beliefs and attitudes. In *Emotions and Bliefs, How feelings influence thoughts.* N H Frijda, A.S.R Manstead and S Bem (Eds). Cambridge University Press, 2000, pp 185, 211.
24. Helm, B. (2001). *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. Cambridge: Cambridge University Press.
25. James, W. (1984): What is an emotion?. *Mind*. 9: 188-205.

26. Jensen, F.V.(2001). ***Bayesian networks and decision graph***. Springer Verlag. 2001.
27. Kahneman, D. Slovic, P. and Tversky,A.(1982): *Judgment under uncertainty, heuristics and biases*. Cambridge University Press, 1982.
28. Izard, C.E.(1993): Four systems for emotion activation: cognitive and non cognitive processes. (1993) Psychological Review 100(1): 68.90.
29. Lazarus R.S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
30. Mc Crae, R. and John, O.P. (1992). An introduction to the Five-Factor model and its applications. Journal of Personality. Vol. 60, pp. 175-215.
31. Miceli, M. & Castelfranchi, C. (1997). Basic principles of psychic suffering: A preliminary account. *Theory & Psychology, 7*, 769-798.
32. Miceli, M. & Castelfranchi, C. (2007). The envious mind. *Cognition and Emotion, 21*, 449-479.
33. Miceli, M., de Rosis, F. and Poggi, I.(2006): Emotional and non emotional persuasion. Applied Artificial Intelligence: an International Journal.
34. Nicholson, A E. and Brady, J. M (1994): Dynamic belief networks for discrete monitoring. IEEE Transactions on Systems, Men and Cybernetics, 24(11), pp. 1593-1610. 1994
35. Oatley, K. and Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. Cognition and Emotion. Vol. 13 pp. 29-50.
36. Ortony, A., Clore, G.L., Collins, A.(1988): The cognitive strusture of emotions. Cambridge University Press
37. Picard, R.W. (1997). *Affective Computing*. The MIT Press.
38. Poggi,I. and Pelachaud,C. (1998). Performative Faces. Speech Communication, 26, pp. 5-21.
39. Poggi I. (2005). The goals of persuasion. *Pragmatics and Cognition*. 13 (2), 297-336.
40. Plutchik, R.(1980): A general psycho-evolutionary theory of emotion. In Plutchik, R. and H. Kellerman, editors, Emotion: theory, research, and experiences (1980), Vol. 1: Theories of emotion, pp. 3.33. Academic.
41. Prendinger, H., Descamps, S. & 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.
42. Prendinger, H. and Ishizuka, M (2005). The empathic companion. A character-based interface that addresses users' affective states. AAI, 19, pp 267-285
43. Reilly, N. (1996). Believable social and emotional agents. Ph.D. diss., School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.
44. Sabini, J. & Silver. M. (1982). *Moralities of everyday life*. New York: Oxford University Press
45. Scherer, K.R., Wranik, T., Sangsue, J., Tran, V., & Scherer, U. (2004). Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction pattern. Soc. Sci. Info.: 43 (4), 499--570.
46. Sloman, S.A.(1996): The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 1, 1996, 3-22
47. Turrini, P., Meyer, J.J., Castelfranchi, C. (2007) Controlling Emotions by Changing Friends. In Dastani, M. & Bordini, R. (eds.) *Proceedings EUMAS'07*, Springer.
48. van Reekum, C. M., & Scherer, K. R. (1997). Levels of processing in emotion-antecedent appraisal. In G. Matthews (Ed.), Cognitive science perspectives on personality and emotion (pp. 259 - 300). Amsterdam: Elsevier Science.