AISB'05 Workshop on 'Mind Minding Agents'

In Favour of Cognitive Models of Emotions

Valeria Carofiglio
Department of Computer Science
Via Orabona 4, 70124 Bari - Italy
carofiglio@di.uniba.it

Fiorella de Rosis
Department of Computer Science
Via Orabona 4, 70124 Bari - Italy
derosis@di.uniba.it

Abstract

We discuss how cognitive models enable integrating recognition of the emotional state with *interpretation of the* reasons of this state and to reason on the potential *impact* of a conversational move on the mental state of the interlocutor. We propose dynamic belief network as a representation formalism for this kind of models.

1 Introduction

Modern theories of emotions recognize that, as soon as we have any experience, we become emotionally aroused to a greater or lesser extent. Factors which may activate emotions are either exogenous (events in the world) or endogenous (internal thoughts and sensations). An example of exogenous stimulus: *When I saw the pictures of the terrorist attack to the Twins Towers I felt shocked and anxious*. Endogenous: *When I imagined the consequences of this attack I felt angry*. In several circumstances, feeling of emotions implies an attempt to interpret them: and interpretation is a cognitive act. Emotional motivations are behind -or rather, before- several intellectual activities. This means that emotion and cognition are inseparable. In human-human dialogs, emotions are transmitted from an interlocutor to the other by mixing and decaying over time and affect their behaviour. Understanding the interlocutor's emotional state may be essential for planning the communicative behavior to adopt in a given context. This is particularly crucial when communication is aimed at suggesting a course of action that, for some reason, the interlocutor may find difficult to follow: typically, cease smoking or change eating habits. In this case, the amount and type of information provided must be calibrated to the attitude of the interlocutor: her knowledge of what a correct behavior is, her belief that her behavior is incorrect, her intention to change it and her definition of a plan to achieve this goal (Prochaska and Di Clemente, 1992).

Knowledge of the cognitive and the emotional state of the interlocutor, combined with the ability to reason about the expected emotional impact of a candidate communicative plan, may therefore allow the speaker to select the best influencing strategy. An advice-giving dialog system which considers the affective aspects of the speaker-user interaction therefore needs a consistent model of the user, which extends the BDI approach with an emotional component (BDI&E). This model enables the system to integrate recognition of the emotional state with *interpretation of the* reasons of this state, according to facts in its knowledge base. This enables ~~it~~ to reason, as well, on the potential *impact* of a conversational move on the mental state of the user. Cognitive models allow achieving these goals: they use principles of cognitive psychology to reason about the link among beliefs, values, goals and activation of emotional states (Castelfranchi,2000). They use psyco-linguistic theories to reason about the relationship between (verbal and non-verbal) expressions and mental states (Poggi and Magno-Caldognetto, 2003). They may employ methods which insure the level of expressivity that is needed to handle partial and uncertain knowledge, dynamic phenomena and variation of effects with the context. In this paper, we propose to introduce cognitive models in persuasive affective dialogs between BDI&E Agents and describe how Dynamic Belief Networks (DBNs) may be employed to represent them (Nicholson and Brady, 1994; Pearl, 2000).

2 Emotionally Oriented Communication (EOC)

After Austin, verbal communication has been seen as involving linguistic 'acts', that is *actions performed by means of words*, originating from a goal and producing a change on the world. When communication is emotionally-oriented, intelligent software agents should be able to plan their (communicative) behaviour by means of an internal mechanism inspired by a consistent combination of cognition and emotion. The inspiration for the agent architecture comes from the recognition that thoughts and feelings are inseparable. The basic *sense-think-act* loop of a BDI agent (Rao and Georgeff, 1991) may be modified to represent the idea that actions are a result of both *thinking* and *feeling*, as shown in figure 1.

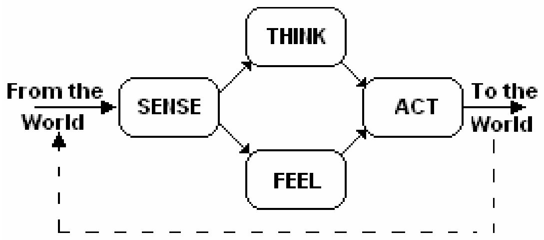

Figure 1: Emotionally Oriented Intelligent Agent Architecture

Representing concepts like *mood*, *emotional state* and *temperament* has been the goal of several research groups. Some of them extended language constructs employed for cognitive modeling to include representation of affective components (Ball, 2002; Bickmore, 2003; Carofiglio et al., in press). However, these systems handle the two components separately. What's interesting, in our view, is to define a framework which enables (i) *to insure consistency between what an agent thinks (the cognitive state) and feels (the emotional state) over time* and (ii) *to exploit this consistent knowledge to plan a communicative act and to interpret the interlocutor's emotional expressions.* In our proposal, the core of this framework is a truth maintenance system which works on *enforcing consistent emotional & rational behavior.* As planning a communicative act requires predicting the interlocutor's behaviour consequent to this act, then *predicting this behavior depends on how this enforcement is carried out*. The agent architecture in figure 1 allows a bi-directional kind of reasoning:

- a *what-if* type of reasoning (direction of the arrows) allows to reason on the *emotional and rational impact* of a communicative act on a given interlocutor starting from some knowledge of her mental state, and therefore to forecast –even if with uncertainty– how this state will be affected by communication;

- a *guessing* type of reasoning (opposite direction of the arrows) allows to: (i) *hypothesize the mental state which possibly produced a 'recognized' emotion* and (ii) *establish the event* (or the events) *which contributed to produce it*, by choosing among several alternative hypotheses.

Our unified framework is employed for several purposes:

(i) *To represent second-order knowledge about the interlocutor's 'mental state',* that we define to be a consistent combination of cognitive and emotional components. A mental state is *valid* (probable, plausible) as long as there is no emotional information to indicate that its cognitive component is inappropriate, and vice-versa.

(ii) *To select a 'convenient' communicative strategy* in a set of alternatives by means of what-if type of reasoning and *to increase the impact of communication, by showing its reasons of validity.*

Let us apply the sense-think-feel-act loop in figure 1 to a simple example about the emotion of fear, in which S denotes the system and U the user. According to the OCC classification (Ortony *et al.*, 1988) and to Oatley and Johnson-Laird's theory (Oatley *et al.*, 1987):

- SENSE corresponds to receiving a communication that a *future, negative event Ev* may occur to U;
- THINK is the combination of three related factors: (i) U's belief that *Ev* will occurr to herself in future; (ii) the value U associates with the goal of *preserving the good of self* and (iii) U's belief that this goal may be threatened;
- FEEL is the emotion of fear;
- ACT consists in showing the fear, through verbal, nonverbal or other kinds of behaviour.

Let us adopt the following notations: A_i, A_h denote the two interlocutors of the dialog; x_i denotes a domain fact; *a* denotes an action; *g* denotes an agent's goal; *e* denotes an emotion. The following atomic formulae stay for respective sentences: *Ev-Has*(A_h, x_i), for "x_i *will occur to* A_h*, sometimes in the future*"; *Ev-Thr*(A_h,*g*) for "*g will be threatened sometimes in the future*"; *Do*(A_h,*a*) for "A_h *performs a*"; *Undesirable*(A_h, x_i) for "x_i *is an undesirable domain state for* A_h" and (*Feel* A_i *e*) for "A_i *feels e*". We call F_i a combination of atomic formulae with $\wedge$,$\vee$, not and $\rightarrow$ connectives, and introduce the goal-formulae (*Goal* A_h F_i) for "A_h *wants that* F_i", the belief-formulae (*Bel* A_h F_i) for "A_h *believes that* F_i" and the communication-formulae (*Say* A_h F_i) for " A_h *says* F_i".

To discuss an example about risks of smoking, we now attribute the following values to the mentioned variables: a=Smoking; x_1=SkinAgeing; x_2=FoetusAtRisk; g=GoodOfSelf. We will then have: *Do*(A_h, *Smoking*) for "*Agent* A_h *smokes*"; (*Ev-Has*(A_h,*FoetusAtRisk*)) for: "A_h*'s foetus will be at risk*"; (*Ev-Has*(A_h,*AgedSkin*)) for "A_h*'s skin will incur an ageing process*"; F_1: (*Do*(A_h,*Smoking*) $\rightarrow$ *Ev-Has*(A_h,*FoetusAtRisk*)) for "*Smoking may produce risks for the foetus*", F_2: (*Do*(A_h,*Smoking*) $\rightarrow$ *Ev-Has*(A_h,*AgedSkin*)) for "*Smoking may produce ageing of skin*", (*Goal* A_h *not Ev-Has*(A_h

AgedSkin)) for *"*A_h *wants to preserve her skin young* ", *(Goal* A_h *not Ev-Has(*A_h *FoetusAtRisk))* for *"*A_h *wants to avoid risks for her newborn"*etc. Let us now see how agent A_i may employ this knowledge to reason about the interlocutor A_h's mind:

(i) *What-if* type of reasoning: we consider the two events:
Ev_1: (Say A_h F_1) and *Ev_2: (Say A_h F_2).*
Which of them will, more likely, activate fear in A_h*?* In selecting a communicative act aimed at convincing A_h to cease smoking, A_i will select between Ev_1 or Ev_2 by considering A_h's beliefs, goals and values (and therefore, her attitudes to 'feel' emotions).

(ii) *Guessing* type of reasoning: *After receiving (from* A_i *or from elsewhere) a message about overall damages of smoking,* A_h *displays signs of fear. Is this fear most likely due to her belief that* F_1 *or that* F_2 *will occur to herself?* If A_i may answer this question, after 'observing' A_h's fear he may exploit knowledge of the *reasons why the communicative act was considered as valid*, to reinforce his persuasive action. For example: "*May be you are afraid of the effects of smoking on your skin: but do consider that cease smoking deletes this effect in a rather short time".*

(iii) *Consistent knowledge* about mental and emotional state: in the example above, if after Ev_1 A_h displays a skeptical expression, A_i may guess that she probably does not believe that "*Smoking may produce ageing of skin*" because this belief is unlikely, given the emotion she displayed. In other circumstances, fear due to the *possibility that the foetus will be at risk* may be unlikely if, for instance, A_i believes that A_h does not want to have a baby.

In the following Section, we will show how DBNs allow us to simulate the described situations. Although, for consistency reasons, we will employ examples based on fear, the method may be applied to any event-based emotion in the OCC classification.

3 Modelling EOC with DBNs.

As we said, tailoring an emotionally oriented advice-giving policy to the attitude of the interlocutor requires some knowledge of her attitude, of alternative persuasion strategies and of strategy-selection criteria. As decision occurs in an evolving and uncertain situation, the process is inherently dynamic. What an agent A_i says is a function of its own state of mind and of its image of the interlocutor A_h's mind. Our analysis will focus on this component and, to simplify the formulae, will omit from second-order beliefs the Bel A_i prefix. We will briefly outline the emotion triggering component that we described extensively elsewhere (Carofiglio et al, in press) by considering, as we said, the example of fear.

Our departure point is that emotions are triggered in A_h by the belief that a particular goal (which is important for the agent) may be achieved or is threatened. So, our simulation is focused on A_i's belief about the *change* in A_h's belief about achievement (or threatening) of her goals over time. We use DBNs as a goal monitoring system that employs the observation data in the time interval (T_i, T_{i+1}) to generate a probabilistic model of the interlocutor's mind at time T_{i+1}, from the model that was built at time T_i (Nicholson and Brady, 1994). Let us consider the triggering of fear that is shown in figure 2 (forget, for the moment, the '+' and '-' labels, whose meaning will become clear later on). The intensity of this emotion in A_h is influenced by the following cognitive components: (i) A_h's belief that x_i will occur to self in the future: *(Bel A_h, Ev-Has(A_h,x_i))*; (ii) the belief that this event is undesirable and therefore A_h does not want it to occur: *(Goal A_h, not Ev-Has(A_h,x_i));* (iii) the consequent belief that this situation may threaten A_h's goal of self-preservation: *(Bel A_h Ev-Thr(A_h,GoodOfSelf)).* Figure 2 shows a compact notation for time-stamped models, Jensen, 2001): the double-arrows indicate temporal links. The number "2" indicates the number of slices. The intensity of the felt emotion depends on the variation of the probability associated with *(Bel A_h Ev-Thr(A_h,GoodOfSelf))* at two consecutive time slices, which is produced when an evidence about some undesirable event is propagated in the network. In our example, this event may either be *(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,AgedSkin)))* or *(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,FoetusAtRisk)))*. It depends, as well, on the weight A_h attaches to achieving that goal, which is a function of the agent's personality. In the mentioned paper, we showed how DBNs enable representing situations that produce emotion mixing due to concurrent triggering of emotions and/or switching among different (and possibly contrasting) emotions.

In addition to the cognitive factors which activate emotion arousal, the model in figure 2 includes other 'rational' components of the state of the mind. According to the Transtheoretical Transaction Theory (Prochaska *et Al.*, 1992), at least three mutually exclusive stages of change may occur in a subject with health behavior problems due to some action a: *Pre-contemplation*, *Contemplation* and *Action*. To represent these stages, we introduce the variable *StageOfChange(Do(A_h,a))* which is influenced by the following cognitive components: (i) A_h's knowledge that she is doing action a: *(Bel A_h W-rong(A_h,a));* (ii) her belief that an event will occur to self in the future as as consequence of doing this action: *(Bel A_h (Do(y,a)→Ev-Has(A_h,x_i)));* (iii) her belief that this event is undesirable: *(Bel A_h Undesirable(A_h ,x_i))*. Due to space limits, we omit from figure 2 the causes of *(Intends A_h Change(A_h,a))* and *(KnowsHow A_h Change(A_h,a))*, wich may be represented by sub-networks similar to the one described for *(Bel A_h Wrong(A_h,a))*. The link between *StageOfChange (Do(A_h,a)) and (Feel A_h Fear)* reflects the fact that the stage of change affects the emotional state, in every time slice.

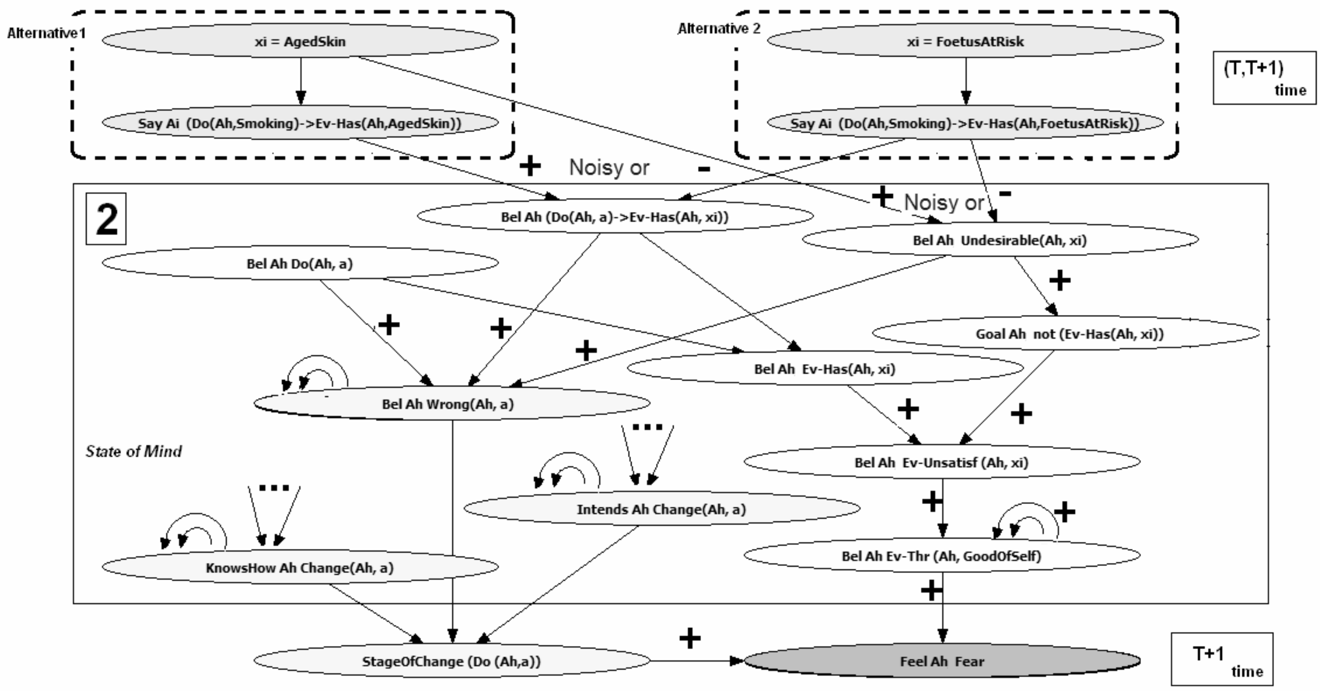

Figure 2: activation of fear

The model in figure 2 contains some hidden assumptions which can be inferred from d-separation properties of BNs. First, it assumes the Markov property: if we know the present, then the past has no influence on the future. In the language of d-separation, the assumption is that *(Bel A_h Ev-Thr(A_h,GoodOfSelf))* at time T-1 is d-separated from the same belief at time T+1 given the belief at time T (and the same for *(KnowsHow A_h Change(A_h, a))*, *(Intends A_h Change(A_h,a))*–and *(Bel A_h Wrong(A_h,a))*. The second hidden assumption has to do with the relationship between stage of change and felt emotion. *StageOf-Change(Do(A_h,a))* and *(Bel A_h Ev-Thr(A_h,GoodOfSelf))* nodes are d-separated, unless some evidence on the node which represents the felt emotion is inserted and propagated in the network. This means that the probability of the stage of change -*StageOfChange(Do(A_h, a))*- is independent of whether there are conditions for an active emotional state. In other words, in figure 2, the fact that A_h may be (for example) in a stage of contemplation according to her belief and goals has no influence on her belief that a given situation may favour threatening her goal of *self-preservation*: *(Bel A_h Ev-Thr(A_h,GoodOfSelf))*. Vice-versa, if en emotion of fear is (directly) observed, that is an evidence about the node representing feeling of this emotion is introduced, *StageOfChange(Do(A_h,a)) and (Bel A_h Ev-Thr(A_h,GoodOfSelf))* become dependent, given *(Feel A_h Fear)*. The model may be employed by A_i to select a persuasive communicative act tailored to A_h by accessing a library of alternatives, all represented as BNs. In principle, every alternative represents a sub-network (see "*Alternative1*" or "*Alternative 2*", in figure 2) which is dynamically 'patched' to the BN representing the image of A_h's mind. If several alternatives related to the same action *a* exist, they are all represented in the network with the method of *noisy functional dependence* (Jensen, 2001): either *Noisy-Or* or *Noisy-And* may be employed to combine alternatives in an appropriate way. For example: in figure 2, "*Alternative1*" and "*Alternative 2*" are combined so that impacts of causes *(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,AgedSkin)))* and *(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,FoetusAtRisk)))* are independent of each other (Noisy-Or).

To investigate the effects of evidence on some alternative *hypotheses*, we employ a qualitative approach, which reduces the problem of parameter estimation (Wellman, 1990). For two generic nodes *A* and *C*, respectively taking states $\{a, \neg a\}$ and $\{c, \neg c\}$, such that $A \rightarrow C$, we say that:

(i) the possibility of C taking value c follows ("+") the possibility of A taking value a if P(c|a) > P(c);
(ii) the possibility of C taking value c varies inversely with ("-") the possibility of A taking value a if P(c|a) < P(c);
(iii) the possibility of C taking value c is independedent of ("0") the possibility of A taking value a if P(c|a)=P(c).

This approach may be applied to forecast the qualitative change in the probability of the hearer *A_h feeling a given emotion, as a consequence of a given communicative act by the speaker A_i*. To answer this question, we observe the qualitative influences among the values of the variables associated with the nodes in the BN in figure 2. Labels '+' and '-' in this figure indicate qualitative dynamic changes in this network, as a consequence of propagating new evidence in it. By means of 'qualitative belief propagation' (Drudzel and Henrion, 1993), we trace the effect of an observation on some node in the BN by propagating the sign of change from the observed node through the entire BN. Every node in the BN, different from the observed one, is given a label which characterizes the sign of the impact of the observed node on the current node.

4 An Example

Let us suppose that A_i wants to persuade A_h to stop smoking and that he knows two alternative ways of doing it: mentioning the consequences of smoking on skin ageing or its possible risks for the foetus. By knowing that A_h is a nice girl who cares for her aspect, A_i assumes that she probably attaches a high weight to avoid ageing of her skin: A_i exploits this knowledge to select the most promising persuasion strategy, by applying a 'what-if' type of reasoning on his model of A_h ; he comes to the conclusion that, if he will say *"Do you know that smoking increases considerably the risk of skin ageing?"*, this will probably induce a fear in A_h and will contribute to persuade her to change of attitude towards smoking. He performs his move and observes A_h's reaction. Now, let us suppose that A_h just says: *"So what?"* without showing any trace of fear. A_h understands that his strategy was not as effective as he expected and tries to 'guess' which might be the reason of this failure. He finds two possible explanations for this: either A_h was not convinced about the association between smoking and skin ageing, or she does not attach much importance to her aspect: in the first case, he might try to employ his argumentation knowledge (for instance, an 'appeal to expert opinion': Walton, 1992) to increase the chance of success of his attempt; in the second one, he might change of strategy by mentioning the risks of smoking for the foetus. Once again, he will monitor the effect of his attempt by observing whether A_h displays any form of concern and will update his model of A_h accordingly.

4.1 Simulating 'what-if' reasoning

In our example, A_i tests, first of all, the effect of an evidence about the node:
(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,AgedSkin))) on the node: *(Feel A_h Fear).*
We set, in figure 2, a=*Smoking*, x_i=*AgedSkin* and the sign of every node to 0, and begin the simulation by sending a positive sign to the evidence node. The node *(Bel A_h (Do(A_h,a)→Ev-Has (A_h,x_i)))* will be updated according to the sign of the link: updating gives sign (+) to this node. Given that *(Bel A_h (Do(A_h,a)→Ev-Has(A_h,x_i)))* is d-connected with *(Bel A_h Do(A_h,a))* and *(Bel A_h Undesirable(A_h ,x_i)),* it sends a message to these nodes. It sends, as well, an indirect positive message to *(Bel A_h Ev-Has(A_h, x_i)).* Analogous reasoning gives sign (+) to *(Feel-A_h Fear).*
At the same time (and with a similar procedure), propagating in the BN an evidence about *(Say A_i (Do(A_h,Smoking)→Ev-Has(A_h,AgedSkin)))* produces a positive change on the node *(Bel A_h Wrong(A_h,a))*. Therefore, A_i anticipates that his communication of the risks of smoking on skin ageing will produce, at the same time, an emotional effect on A_h and a change in her belief that she is adopting a wrong behaviour. This change may be slight or large, depending on A_h characteristics and also on the context in which communication occurs: the final result may be a change from the 'precontemplation' to the 'contemplation' stage, which requires (to the system) an adequate change of advice-giving strategy.

4.2 Simulating 'guessing' reasoning

Let us go on with our example, by considering what happens after A_h says *"So what?"*without expressing any concern. To understand the possible rasons of his failure, A_i reasons on the *most probable configuration* of this fact, that is the *most probable explanation* of this evidence (Pearl, 2000). As in 'what if' reasoning, this may be achieved by reasoning on the qualitative influence among the variables associated with the nodes in the BN in figure 2. A negative value of the 'fear' node, together with a positive value of the *(Bel A_h, Do(A_h,Smoking))* node, produce a negative value for the nodes *(Bel A_h Ev-Has(A_h,x_i))* and *(Bel A_h Undesirable(A_h,x_i)).* These are two possible explanations of the move failure that A_i will try to repair.

5 Conclusions

There may be at least two objections to our modelling method. The first one is the always raised question of ‘where are the parameters in the model coming from’. In cognitive models, parameters cannot be learned by knowledge discovery methods, as a dataset including observations about ‘states of mind’ is hard to get. Subjective estimate is therefore the only reasonable procedure to apply. To reduce the risk of errors in these estimates we adopt, as we said, a qualitative approach to reasoning which does not pretend to measure exactly the changes introduced in the various nodes by new evidence acquired but only estimates them qualitatively. On the other side, we make a sensitivity analysis on the model (Jensen, 2001) which enables us to estimate the parameters which mainly contribute to affect the results: this analysis suggests where to focus the parameter estimation work.

The second, and more intriguing, possible objection we may anticipate concerns the hypothesis of consistency between the emotional and the cognitive components of an agent’s state of mind, which (we admit it) is very strong. Emotions do not always (and not immediately) entail consistent reasoning about their reasons: I may feel shocked in saying the pictures of the Twin Towers even without reflecting on this episode for some time. In some cases, one may even claim that reasoning produced by an emotional state might be inconsistent with it (at least apparently). In spite of these limits, we could experience the advantages of our model in simulating used-adapted advice giving dialogs and hope that they might prove to be useful, as well, as a tool for fostering discussion with cognitive psychologists about the mechanisms which govern emotional states.

References

Ball, E.: “A Bayesian Heart: Computer Recognition and Simulation” in "Emotions in Humans and Artifacts". Robert Trappl, Paolo Petta and Sabine Payr (Eds.), MIT Press. 2002

Bickmore T. *Relational Agents: Effecting change through human-computer relationships*. PhD Thesis. 2003

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

Castelfranchi, C. Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions. In A. Paiva, Ed. *Affective Interactions*. Springer LNAI 1814, Berlin:76-106. 2000.

Drudzel, M.J. and Henrion,M.: “Intercausal Reasoning with uninstantied ancestor nodes”. In Proceedings of The Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93), pp 3-11, Washington. 1993

Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer. 2001.

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

Ortony, A., Clore, G.L. and Collins, A. The cognitive structure of emotions. Cambridge University Press, Cambridge, MA. 1988.

Oatley, K. and Johnson-Laird. “Towards a cognitive theory of emotions”. Cognition and Emotion. Vol. 13 pp. 29-50. 1987

Pearl, J.A.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufman Publishers, 2000.

Poggi, I., Magno-Caldognetto, E., “Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici.” Proceedings of the Conference “Il Parlato italiano”. D’Auria, Naples, 2003

Prochaska, J., Di Clemente, C., Norcross, H.: “In search of how people change: applications to addictive behaviour”. Americal Psychologist, 47 (1992) 1102-1114

Rao, A.S. and Georgeff, M.P.: “Modeling rational agents within a BDI-architecture”. In J. Allen, R. Fikes & R. Sandewall (Eds.). Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR’91). Morgan Kaufman. pp. 473-484. 1991

Walton, D. The place of emotion in argument. The Pennsylvania State University Press, 1992.

Wellman, M.P. “Fundamental Concepts of Qualitative probabilistic Networks”. Artificial Intelligence, 44(3): 257-303. 1990

Acknowledgements

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