Cognitive Emotion Models:

some thoughts on their tuning and their possible use in communication processes

V. Carofiglio and F de Rosis

Intelligent Interfaces, Department of Informatics

University of Bari, Italy

{carofiglio,derosis}@di.uniba.it

1. Introduction

In planning a communicative act, intelligent software agents can no longer be modelled as pure goal-driven, task-solving agents: they also have affective reasons for their choices and behaviour which drive decision-making. Hence, they should be able to exhibit a behaviour inspired by a consistent combination of cognition and emotion. 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. This is particularly true 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, information provided must be calibrated to the attitude of the interlocutor: his knowledge of what a correct behaviour is, hsi belief that his behaviour is incorrect, his intention to change it and his definition of a plan to achieve this goal (Prochaska and Di Clemente, 1992). Knowledge about the cognitive and (also) the emotional state of the interlocutor, combined with the ability to reason about the expected emotional impact of a candidate communicative act, may allow the speaker to select the best influencing strategy. In the remainder of this paper, we propose a framework which enables (i) insuring consistency between what an agent thinks (the cognitive state) and feels (the emotional state) over time and (ii) exploiting this consistent knowledge to plan a communicative act. We argue in favor of dynamic belief networks as a formalism to handle this kind of models and discuss how these models may be tuned up.

2. Mental models as combinations of cognition and affect

Our framework includes second-order knowledge about the interlocutor's mental state, that we define to be a consistent combination of cognitive and emotional components. A given cognitive configuration is said to be *valid* (probable, plausible) in a given context as long as there is no emotional information to indicate that this component is inappropriate in that context, and vice-versa. The core of our framework is a truth maintenance system which works on enforcing consistent emotional and rational behaviour. We document this enforcement by showing two different kinds of reasoning with which the framework is endowed:

- a *what-if type of reasoning* 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* 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.

The framework may be employed to *select a 'convenient' communicative strategy* from a set of alternatives by means of a *what-if type of reasoning* and *to increase the impact of communication* by showing its reasons of validity, by means of a *guessing type of reasoning*.

Let A_i, A_h be variables denoting the two interlocutors of the dialog; x_i , a , g and em be variables denoting (respectively) an event, an action, an agent's goal and an emotional state. The following formulae stay for respective sentences:
$Ev\text{-}Has(A_h, x_i)$, for "*x_i will occur to A_h, sometimes in the future*";
$Ev\text{-}Thr(A_h, g)$ 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 undesirable for A_h*" and $(Feel\ A_h, em)$ for "*A_h feels em*".
We call F_i a combination of the above formulae with "→" and "¬"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 that F_i*".

Following the Nicholson and Brady's approach (Nicholson and Brady, 1994), we showed in (de Rosis et Al., 2003) how we used Dynamic Belief Networks (DBN) as a goal monitoring system that employs the observation data in the time interval (T, T+1) to generate a probabilistic model of the agent's mind at time T+1, from the model that was built at time T. We employ this model to reason about the consequences on the monitored goals of the observed event (any change happening in the virtual world, including a communicative act by A_i). Let us consider, as an example, the triggering of fear that is shown in figure 1[1].

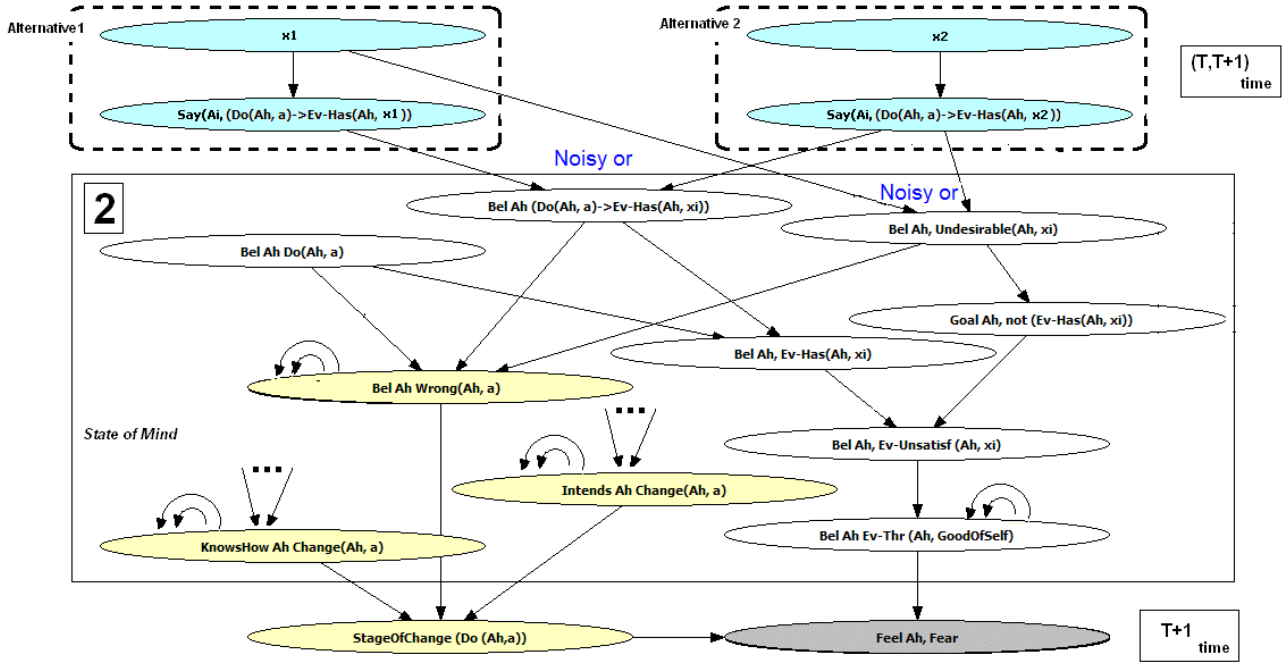

Fig1. The triggering of fear

Fear is a 'negative' emotion which is triggered by a change in the belief that the goal of *Preserving-self-from-future-bad* is threatened. The intensity of this emotion is influenced by the following cognitive components:

- the agent's belief that an event x_i will occur to self in the future *(Bel* A_h *Ev-Has(*A_h*,*x_i*))=TRUE*;
- the belief that this event is undesirable: *(Bel* A_h *Undesirable(*A_h*,*x_i*))=TRUE*;
- the belief that this situation threats the agent's goal of *Preserving-self-from-future-bad*: *(Bel* A_h *Ev-Thr(GoodOfSelf))=True.*

According to utility theory, *the variation in the intensity of an emotion* is a function of the product of the change in the probability to achieve a given goal, times the utility that achieving this goal takes to the agent.

An agent A_i may employ this knowledge to select a communicative act tailored to its interlocutor A_h. Let us consider again fig.1. A_h *applies a what-if type of reasoning (direction of the arrows) to wonder which consequence of action a should best be mentioned, to activate an emotion of fear in* A_h. It considers talking about x_1 *[(Say* A_i *Do(*A_h*,a)→ Ev-Has(*A_h*,*x_1*))]* or x_2 *[(Say* A_i *Do(*A_h*,a)→ Ev-Has(*A_h*,*x_2*))]* and reasons about the impact of the two communicative acts on A_h's mind by considering her beliefs, goals and values and therefore her attitudes to 'feel' emotions

Agent A_i may employ the model to perform, as well, a *guessing-type of reasoning* (opposite direction of arrows) to wonder whether an emotion of fear displayed by A_h *may be due to her belief that* x_1 *or* x_2 will occur to self. In this case, A_i may exploit his knowledge of the *reasons why he considered his communicative act to be potentially valid*, to reinforce his persuasive action. For example: *"You are probably afraid of the effects of a: but do consider that, if you stop making a, these effects will be deleted in a short time"*.

In the context above, the assumption of consistency of thinking and feeling entails that if, after communication about x_1, A_h displays a sceptical expression, A_i may infer that she probably does not believe that x_1 will occur to self, because this belief is unlikely given the emotion she displayed.

[1] Notice that the network includes two symmetrical components (at times T and T+1). Due to limitation in space, we show in the figure only the component at time T+1, with the link with the monitored goal at time T.

3. Tuning the models

A common question to this kind of models is the following: where are the parameters coming from? In cognitive models, parameters cannot be learned with knowledge discovery methods, as a dataset including observations about states of mind is hard to get. In contrast, totally subjective estimation risks to be unreliable. To reduce the risk of errors in these estimations, a reasonable procedure to apply consists in working concurrently with several models by comparing their predictive value. Without going into details, usually in a candidate model a set of questionable links exists. For each of those links, we investigate the robustness of the network which is obtained from the original one after revising the strength of that link. That is, we investigate the effects of imprecision of parameters in the model on the output of the model itself by identifying, every time, the parameters that independently or jointly produce a significant effect on the probability of a node of interest. Notice that these calculations are easy to perform on several models because these models are very similar to each other.

Based on reasoning algorithms for probabilistic inference, efficient computation methods have been developed for this kind of sensitivity analysis. Two aapproaches exist: theoretical and empirical. The theoretical approach establishes a function expressing a posterior probability of interest in terms of the parameters under study (e.g., Laskey, 1995, Coupè et al., 2000). The empirical methods examine the effects of varying the network's parameters on diagnostic or predictive performance of the model(e.g., Pradhan et al., 1996). We briefly show an example of our arguments related to our theoretical approach to sensitivity analysis.

Let us consider again the triggering of fear that was considered in figure 1. We are interested in analysing several aspects of robustness of the model: (i) with respect to the context (e.g is there a friendship relation between A_i and A_h?), (ii) with respect to the personality of A_h (e.g. her attitude towards emotion feeling); (iii) with respect to the desirability of x_i (to A_h or A_i). As a consequence, our evaluation involves different portions of the current model, one at time.

Following the approach in (Coupè et Al., 2000), we assume that the value `True` is observed in [T,T+1] for the variable $((Say\ A_i\ Do(A_h,a)) \rightarrow Ev\text{-}Has(A_h,x_1))$. If, for example, we wish to evaluate the robustness of the network with respect to A_h's attitude to feel emotions, then the probability of interest is the probability that the variable $(Bel\ A_h\ Ev\text{-}Thr(GoodOfSelf))$ at time T+1 takes the value `True`. The sensitivity of:

$$Pr((Bel\ A_h\ Ev\text{-}Thr(GoodOfSelf)|((Say\ A_i\ Do(A_h,a)) \rightarrow Ev\text{-}Has(A_h,x_1))) =$$

$$Pr((Bel\ A_h\ Ev\text{-}Thr(GoodOfSelf) \wedge ((Say\ A_i\ Do(A_h,a) \rightarrow Ev\text{-}Has(A_h\ x_1)))/$$

$$Pr((Say\ A_i\ Do(A_h,a) \rightarrow Ev\text{-}Has(A_h,x_1))) \qquad \textbf{(1)}$$

to variations in the estimates of the network parameter: $P((Bel\ A_h\ Ev\text{-}Thr(GoodOfSelf))_{T+1}|\ (Bel\ A_h\ Ev\text{-}Thr(GoodOfSelf))_T,\ (Bel\ A_h\ Ev\text{-}Unsatisfied(Ah,x_i)))$ is analysed. Without going into numerical details, we need to establish the values of coefficients for the multilinear relations at the numerator and the denominator of equation **(1)**. So far, we have performed only some initial tests. It is our plan to explore the aspects we mentioned in our cognitive models of emotions, in which the process of activation of several mixing emotions with different intensities is represented.

Main References

Coupè, V.M.H., Van der Gaag, L.C. and Habbema, J.D.F. Sensitivity analysis: an aid for belief network quantification. *The Knowledge Engineering Review,* 15 (3), 2000.

de Rosis, F, Pelachaud, C. Poggi,I., Carofiglio,V. and De Carolis,B.. 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. BR Gaines Vol.59 (1-2) 2003*.

Laskey, K.B. Sensitivity analysis for probability assessments in Bayesian Networks. IEEE Transaction on Systems, Men and Cybernetics, 25(6):901-909

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

Pradhan,M.Henrion,M.Provan.G, drlFavero,B. Huang,K. The sensitivity of Belief networks to imprecise probabilities: An experimental investigation. *Artificial Inteligence 85(1-2): 363-397*

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)