Chapter 1

COOPERATIVE INTERFACE AGENTS

Sebastiano Pizzutilo, Berardina De Carolis and Fiorella de Rosis
Department of Computer Science - University of Bari, Italy
{pizzutilo, decarolis, derosis}@di.uniba.it

Abstract Animated Agents are endowed with personality and emotions, with the aim of increasing their believability and of establishing an empathic relationship with the User. In this paper, we claim that, to endow Agents with social intelligence, the 'communication' traits described in the Five Factor Model should be integrated with some 'cooperation' attitudes; we describe our experience in building an Agent that combines the two personality aspects and discuss the problems still open.

Keywords: cooperation, personality-rich agents

1. Introduction

In the near future, computers will either 'disappear', to ubiquitously pervade life environment in a not immediately perceivable way, or will take the appearance of a human being, to undertake a friendship relationship with the User. In both cases, endowing agents with some form of social intelligence appears to be a crucial need. If reasoning, help and control abilities are distributed among specialised agents integrated with objects of daily life, some form of communication and cooperation among them will be needed, to avoid conflicting behaviours. On the other side, a 'believable' embodied agent should be able to understand the Users, help them in solving problems, find ways of coming to a mediated solution in case of conflicts and so on. To enable the Users to foresee how the Agent will behave, it should harmonise its external appearance with its internal behaviour, understand how to adapt to their needs and moods and, finally, enable them to 'select a different partner' if they wish.

Short and long-term variations in the behaviour of Embodied Agents have been metaphorically represented, respectively, in terms of *emotional states* and *personality traits*. Endowing Socially Intelligent Agents with a personality requires defining: (i) which forms of social intelligence these Agents should have and how they may be translated in terms of personality traits; (ii) how a trait may be represented in the Agent's mental state and reasoning style; (iii) how various traits may be combined in the same individual and, finally, (iv) how one or more traits may be manifested in the Agent's behaviour. In this paper, we discuss our experience in building an Interface Agent that cooperates with the User in performing software application tasks; we will focus our description on the way that we formalised its cooperation attitude.

2. Cooperation personalities in XDM-Agent

Research on personality-based HCI has be driven by results of studies about human intelligence: in particular, the 'Five Factor Model' (FFM) and the 'Interpersonal Circumplex Model' (IC). The FFM derives from the psychologists' need of defining '*the most important ways in which individuals differ in their enduring emotional, interpersonal, experiential, attitudinal and motivational styles*' (McCrae and John, 1992). The five dimensions are an interpretation of results of applying factor analysis to questionnaires submitted to various groups of subjects; their meaning is a subjective interpretation of the set of variables they 'explain', and is described with natural language terms. 'Sociability' or 'Social closeness' is associated, in particular, with Extraversion. The second method employed to categorise human personalities is Wiggin's measure of IC, whose axes are 'Dominance' and 'Affiliation'. Whether the two factorisation criteria are related is not fully clear: some authors identify Extraversion with Dominance, while others argue that Extraversion is best seen as located midway between Dominance and Warmth (McCrae and John, 1992).

Researchers in HCI employed the two mentioned factorisation criteria to enrich interfaces with a personality. Some notable examples: Nass and colleagues studied graphical interfaces in terms of Dominance and Agent-based interfaces in terms of Extraversion (Nass *et al.* 1995); Dryer adopted the IC model (1998); Andre' et al (2000) attach Extraversion and Agreableness to the members of their 'Presentation Teams'; Ball and Breese (2000) included Dominance and Friendliness in their modeling of personality-related observable behavior. To computer scientists, the advantage of referring to the two mentioned models is to have a widely accepted frame of reference, with a definition of the way that

every personality factor manifests itself in the external behavior. The main disadvantage is that these personality traits refer to a characterisation of 'communication styles' rather than to mental social attitudes. They are therefore very useful for endowing Agents with a 'pleasant' and 'believable' appearance, but not to express diversification in social relationships. Another difficulty in employing the cited models is that traits are defined through natural language descriptions and are not easily formalised into the 'mental state' of an Agent. The first and most relevant contribution to a cognitive theory of personalities was due to Carbonell (1980), who saw them as combinations of degrees of importance assigned to goals. A second example, to which we will refer in particular in this paper, is Castelfranchi and Falcone's (1998) theory of cooperation in multi-agent systems.

Although affective expressions may contribute to increase the Interface Agent's friendliness, its acceptability is driven by the level of help provided to the User, that is by its 'cooperation attitude'. This *level of help* should not be equal to all Users but should be tailored to their attitudes towards computers in general and towards the specific software to which the Agent is applied in particular. These attitudes may be synthetized in a *level of delegation* of tasks that the User adopts towards the Agent. To select the helping attitude that best suits to the User needs, the Agent has to be endowed with a reasoning capacity that enables it to observe the User, to model her expected abilities and needs and to plan the 'best' response in every context. We had already applied the theory of Castelfranchi and Falcone to formalise the mental state of agents and their reasoning capacities in a previous Project (GOLEM)(Castefranchi *et al.* 1998). With the Project described in this paper, we extended that research in the direction of Embodied Animated Agents.

XDM-Agent is an Embodied Animated Character that helps the User in performing the tasks of a given application: its cooperation attitude changes according to the User and the context. Although the Agent is domain-independent, we will take electronic mail as a case study, to show some examples of how it behaves in helping to use Eudora. In a software of wide use like this, all procedures should be very natural and easy to perform; the first goal of XDM-Agent is then "*to make sure that the User performs the main tasks without too much effort*". At the same time, the Agent should avoid providing too much help when this is not needed; a second goal is therefore "*to make sure that the User does not see the Agent as too much intrusive or ennoying*". These general goals may specialise into more specific ones, according to the 'cooperation attitude' of the Agent. In deciding the level and the type of help to provide, XDM-Agent should consider, at the same time, the

User experience and her 'delegation attitude'. The Agent's decision of whether and how to help the User relies on the following knowledge sources:

Own Mental State, in which the Agent's goals (Goal XDM (T g)) and abilities (Bel XDM (CanDo XDM a)) and the actions it intends to perform (Bel XDM (IntToDo XDM a)) are represented;

Domain Knowledge: XDM should know all the plans that enable achieving tasks in the application: $\forall g \forall p$ (Domain-Goal g)$\wedge$(Domain-Plan p)$\wedge$(Achieves p g) $\Rightarrow$

(KnowAbout XDM g)$\wedge$(KnowAbout XDM p)$\wedge$ (Know XDM (Achieves p g)).

It should know, as well, the individual steps of every domain-plan: $\forall g \forall a$ (Domain-Goal p)$\wedge$(Domain-action a)$\wedge$(Step a p) $\Rightarrow$(KnowAbout XDM p)$\wedge$(KnowAbout XDM a)$\wedge$(Know XDM (Step a p).

User Model: the agent should have some hypothesis about: *i*) the User goals, both in general and in specific phases of interaction [$\forall g$(Goal U (T g)) $\Rightarrow$(Bel XDM (Goal U (Tg)))]; *ii*) her abilities [$\forall a$ (CanDo U a) $\Rightarrow$(Bel XDM (CanDo U a))] and *iii*) what the User expects the Agent to do, in every phase of interaction [$\forall a$ (Goal U (IntToDo XDM a)) $\Rightarrow$(Bel XDM Goal U (IntToDo XDM a))]. This may be a default, stereotypical knowledge about the User that is settled at the beginning of interaction; ideally, the model should be updated dynamically, through plan-recognition.

Reasoning Rules: the Agent employs this knowledge to take decisions about the level of help to provide in any phase of interaction, according to its 'helping attitude', which is represented as a set of 'reasoning rules'. For instance, if XDM-Agent is a *benevolent*, it will respond to all the User's (implicit or explicit) requests of performing actions that it presumes she is not able to do:

Rule R1 $\forall a$[(Bel XDM (Goal U (IntToDo XDM a)))$\wedge$(Bel XDM$\neg$ (CanDo U a))$\wedge$(Bel XDM (CanDo XDM a))] $\Rightarrow$(Bel XDM (IntToDo XDM a)). If, on the contrary, the Agent is a *supplier*, it will do the requested action only if this does not conflict with its own goals:

Rule R2 $\forall a$ [(Bel XDM (Goal U (IntToDo XDM a)))$\wedge$ (Bel XDM (CanDo XDM a)) $\wedge$ ($\neg$ $\exists$ g (Goal XDM (Tg) $\wedge$ (Bel XDM (Conflicts a g))] $\Rightarrow$
(Bel XDM (IntToDo XDM a))

... and so on for the other personality traits.

Let us assume that our Agent is a benevolent and that the domain goal g is "to write a correct email address". In deciding whether to help the User, it will have to check, first of all, how the goal g may be achieved. Let us assume that no conflict exists between g and the Agent's goals. By applying rule R1, XDM will come to the decision to do its best to help the user in writing the address, by directly performing all the steps of the plan. The Agent might select, instead, a *level of help* to provide to the User: this level of help may be seen, as well, as a personality trait. If, for instance, XDM-Agent is a *literal helper*, it will only check that the address is correct. If, on the contrary, it is an *overhelper*, it will go beyond the User request of help to hypothesize her higher order goal (for instance, to be helped in correcting the address, if possible). A *subhelper* will only send a generic error message: this is what Eudora does at present if the User tries to send a message without specifying any

address. If, finally, the User asks the Agent to suggest how to correct the string and the Agent is not able to perform this action and is a *critical helper*, it will select and apply, instead, another plan it knows.

3. Personality Traits' Combination

In multiagent cooperation, an Agent may find itself in the position of delegating some task or helping other Agents. A theory is therefore needed to establish how delegation and helping attitudes may combine in the same Agent. Some general thoughts about this topic may be found in (Castefranchi *et al.* 1998). In XDM-Agent, the Agent's reasoning on whether to help the User ends up with an intentional state (to perform an individual action, an entire plan or part of a plan). This intentional state is transformed into an action that may include communication with the User: for instance, an *overhelper* Agent will interact with the User to specify the error included in the string, will propose alternatives on how the string might be corrected and will ask the User to correct it. In this phase, the Agent will adopt a communication personality trait: for instance, it might do it in an 'extroverted' or an 'introverted' way. The question then is: *how should cooperation and communication personalities be combined?* Is it more reasonable to assume that a overhelper is an extroverted or an introverted? We do not have, at present, an answer to this question. In the present prototype, we implemented only two personalities (a *benevolent* and a *supplier*) and we associated the 'benevolent' trait with the 'extroverted' one and the 'supplier' with the 'introverted'.

The User desire to receive help may be formalised, as well, in personality terms: if the User is a '*lazy*', she expects to receive, from XDM, some cooperation in completing a task, even if she would be able to do it by herself (and therefore, irrespectively of her 'level of experience').

Rule R3: $\forall a \forall g$[(Goal U (T g))$\wedge$(Bel U (Achieves ag))$\wedge$(Bel XDM (CanDo XDM a)) $\Rightarrow$(Goal U (IntToDo XDM a))].

If, on the contrary, the User is a '*delegating-if-needed*', she will need help only if she is not able to do the job by herself (for instance, if she is a 'novice'):

Rule R4: $\forall a \forall g$ [(Goal U (T g))$\wedge$(Bel U (Achieves a g))$\wedge$(Bel XDM $\neg$(CanDo U a))$\wedge$(Bel XDM (CanDo XDM a)) $\Rightarrow$(Goal U (IntToDo XDM a))].

Providing help to an 'expert' and 'delegating-if-needed' User will be seen as a kind of intrusiveness, that will violate the Agent's goal to 'avoid annoying the User'.

In our first prototype of XDM-Agent, the Agent's cooperation personality (and therefore its helping behaviour) may be settled by the User

at the beginning of interaction or may be selected according to some hypothesis about the User. As we said before, the Agent should be endowed with a plan recognition ability that enables it to update dynamically its image of the User: notice that, while recognising communication traits requires observing the 'external' (verbal and nonverbal) behavior of the User, inferring the cooperation attitude requires reasoning on the history of interaction (a 'cognitive diagnosis' task that we studied, in probabilistic terms, in (de Rosis *et al.* in press)). Once some hypothesis about the User's delegation personality exists, how should the Agent's helping personality be settled? One of the controversial results of research about communication personalities in HCI is whether the similarity or the complementarity principles hold: that is, whether an 'extroverted' interface should be proposed to an 'extroverted' User, or the contrary. When cooperation personalities are considered, the question becomes the following: "How much should an Interface Agent help a User? How much importance should be given to the User experience (and therefore her abilities in performing a given task), how much to her propensity to delegate that task?". In our opinion, the answer to this question is not unique. If XDM-Agent's goals are those mentioned before, that is "*to make sure that the User performs the main tasks without too much effort*" and "*to make sure that the User does not see the Agent as too much intrusive or ennoying*", then the following combination rules may be adopted:

(DelegatingIfNeeded U)⇒(Benevolent XDM): the Agent helps delegating-if-needed Users only if it presumes that they cannot do the action by themselves;

(Lazy U)⇒(Supplier XDM): theAgent does its best to help lazy users, unless this conflicts with its own goals;

.... and so on. However, if the Agent has also the goal *"to make sure that Users exercise their abilities"* (such as in Tutoring Systems), then the matching criteria will be different; for instance:

(Lazy U)⇒(Benevolent XDM): the Agent helps a lazy User only after checking that she is not able to do the job by herself. In this case, the Agent's cooperation behavior will be combined with a communication behavior (for instance, Agreableness) that warmly encourages the User in trying to solve the problem by herself.

XDM-Agent has been implemented by trying to achieve a distinction between its external appearance (its 'Body', developed with MS-Agent) and its internal behaviour (its 'Mind', developed in Java). It appears as a character that can take several bodies, can move on the display to indicate objects and make several other gestures, can speak and write a text in a balloon. To insure that its body is consistent with its mind, the ideal would be to match the Agent's appearance with its helping personality: however, as we said, no data are available on how cooperation

traits manifest themselves, while literature is rich on how communication traits are externalised. At present, therefore, XDM-Agent's body only depends on its communication personality; we associate a different character with each of them (*Genie* with the benevolent-extroverted and *Robby* with the supplier-introverted). However, MS-Agent enables us to program the Agent to perform a minimal part of the gestures we would need: we are therefore working, at the same time, at a more refined Animated Agent that can adapt its face, mouth and gaze to its high-level goals, beliefs and emotional states: this will enable us to directly link individual components of the Agent's mind to its verbal and non-verbal behaviour, through a set of personality-related activation rules (Poggi *et al.* in press).

4. Conclusions

Animated Agents tend to be endowed with a personality and with the possibility to feel and display emotions, for several reasons. In Tutoring Systems, display of emotions enables the Agent to show to the students that it cares about them and is sensitive to their emotions; it favours convey of enthusiasm and contributes to insure that the student enjoys during learning (Elliott *et al.* 1997). In Information- Providing Systems, personality traits contribute to specify a motivational profile of the Agent and to orient the dialog accordingly (Andre' *et al.* 2000). Personality and emotions are attached to Personal Service Assistants to better 'anthropomorphize' them (Arafa *et al.* 1998). As we said at the beginning of this paper, personality traits that are attached to the Agents reproduce the Big-Five Factors that seem to characterise human social relations. Among the traits that have been considered so far, 'Dominance/Submissiveness' is the only one that relates to cooperation attitudes: according to Nass and colleagues, 'Dominants' are those who pretend that others help them when they need it; at the same time, they tend to help others by assuming responsibilities on themselves. 'Submissives', on the contrary, tend, on one side, to obey to orders and, on the other side, to delegate actions and responsibilities whenever possible. This model seems, however, to consider only some combinations of cooperation and communication attitudes that need, on the contrary, to be studied and modeled separately and more indepth. We claim that Castelfranchi and Falcone's theory of cooperation might contribute to such a goal, and the first results obtained with our XDM-Agent prototype encourage us to go on in this direction. As we said, however, much work has still to be done to understand how psychologically plausible

configurations of traits may be defined, how they evolve dynamically during interaction and how they are externalised.

References

Andre', E., Rist, T.:, van Mulken, S., Klesen, M. and Baldes, S. (2000). The automated design of believable dialogues for animated presentation teams. In Cassel, J, Sullivan, J., Prevost, S. and Churchill, E. (Eds). *Embodied Conversational Agents*. The MIT Press.

Arafa, J., Charlton, P., Mamdani, A. and Fehin, P. (1998). Designing and building Personal Service Assistants with personality. In Prevost, S. and Churchill, E. (Eds). *Proceedings of the Workshop on Embodied Conversational Characters*.

Ball, G. and Breese, J. (2000). Emotion and Personality in a Conversational Agent. In Cassel, J., Sullivan, J., Prevost, S. and Churchill, E. (Eds). *Embodied Conversational Agents*: The MIT Press.

Carbonell, J. (1980). Towards a process model of human personality traits. *Artificial Intelligence*, 15.

Castelfranchi, C. and Falcone, R. (1998). Towards a theory of delegation for agent-based systems. *Robotics and Autonomous Systems, Special Issue on Multi-Agent Rationality*.

Castelfranchi, C., de Rosis, F., Falcone, R. and Pizzutilo, S. (1998). Personality traits and social attitudes in multiagent cooperation. *Applied Artificial Intelligence*, 12: 7-8.

de Rosis, F., Covino, E., Falcone, R., and Castelfranchi, C. (in press). Bayesian cognitive diagnosis in believable multiagent systems. In: Williams, M. A. and Rott, H. (Eds), *Frontiers of Belief Revision*. 'Applied Logic' Series of Kluwer.

Dryer D.C. (1998). Dominance and valence: a two-factor model for emotion in HCI. *Proceedings of AAAI*.

Elliott, C., Lester, J.C. and Rickel, J. (1997). Interpreting affective computing into animated tutoring agents. IJCAI Workshop on *Intelligent Interface Agents*.

McCrae, R. R. and John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B. and Dryer, D.C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43.

Poggi, I., Pelachaud, C. and de Rosis, F. (in press). Eye communication in a conversational 3D synthetic agent. *AI Communications*.