Personality-Driven Social Behaviors in Believable Agents

Paola Rizzo
Center of Cognitive Science
University of Torino
Via Lagrange 3
I-10123, Torino, Italy

Manuela Veloso
Carnegie Mellon University
Computer Science Dpt.
5000 Forbes Avenue
Pittsburgh, PA 15213-3891, USA

Maria Miceli
IP-CNR
National Research Council
Viale Marx 15
I-00137, Roma, Italy

Amedeo Cesta*
IP-CNR
National Research Council
Viale Marx 15
I-00137, Roma, Italy

Abstract

Agents are considered "believable" when they are viewed by an audience as endowed with thoughts, desires, and emotions, typical of different personalities. The paper describes our work in progress aimed at realizing believable agents that perform helping behaviors influenced by their own personalities; the latter are represented as different clusters of prioritized goals and preferences over plans for achieving the goals. The implementation is based on the integration of a state-based planner that serves as the reasoning tool for the agents and a situation-driven execution system.

Introduction

"There is a notion in the Arts of 'believable character'. It does not mean an honest or reliable character, but one that provides the illusion of life, and thus permits the audience's suspension of disbelief[1]*. The idea of believability has long been studied and explored in literature, theater, film, radio, drama, and other media".* (Bates 1994) Believability therefore refers to a character's features producing the illusion of an "animated" being, able to think, feel emotions, and having attitudes toward others, etc., in a way very similar to a human being.

Recently several research projects and groups (the Oz project at CMU (Bates 1994), the Agents group at MIT (Maes 1995), the Virtual Theater project at Stanford (Hayes-Roth, Brownston, & vanGent 1995), etc.) have focused their efforts on the creation of computerized believable agents (BAs). The important novelty of computer-based BAs compared with those created with traditional media concerns their *interactivity*: in fact, while it is only possible to passively witness the actions performed by movies' or cartoons' characters, with interactive BAs one can talk, play, fight, and so on. The frontiers of virtual reality make this possibility even more interesting. The applications of BAs concern mainly human-computer interaction, interactive entertainment, education, and some artistic domains.

The realization of computer-based interactive BAs presents problems that are different from those typical of traditional media. In fact, while movies and cartoons are designed and realized off-line, computer-based BAs must be able to interact with a human user in many kinds of situation; therefore, their behavior cannot be a priori defined, but it must be as autonomous and flexible as possible.

From the relevant literature it emerges that among the most important features of a BA are a marked personality and emotions; therefore, most works on BAs are focused on the physical and behavioral features more suited to express emotions, and study how to realize the personalities of BAs by varying the mapping between emotions and behaviors (e.g., if the user threatens a BA, the latter will be scared and will flight if it is a fearful agents, while it will become angry and will attack the user if it is a brave agent), or by varying the "traits" or features underlying personality.

In this work the analysis and creation of goals and preferences underlying agents' personalities takes priority over the realization of characteristic emotional reactions and over trait-based representations with the purpose of realizing believable agents endowed with more complex personalities and behaviors.

The paper is structured as follows. First, we support our view of a goal-based approach to personality and present our model. We instantiate our model of personality by describing the goals which influence the behavior of help-giving, and briefly describe the implementation under realization with a planner and a reactive execution system. Finally, we draw some conclusions.

A goal-based model of personality

There are several approaches at realizing believable agents with markedly different personalities. Given the importance attributed to emotions for BAs, personalities are often constructed by varying the mapping between emotions and reactions. In other words, each emotion can be associated with a class of reac-

* e-mail: paola@di.unito.it, mmv@cs.cmu.edu, {maria, amedeo}@pscs2.irmkant.rm.cnr.it

[1] The concept of "suspension of disbelief" comes from the theater literature; for clarifications, see (Laurel 1991)

tions typical of each personality: for example, fear can be associated with a class of flight reactions for realizing a fearful personality, or to aggressive behaviors in order to realize an aggressive personality (Bates 1994; Bates, Loyall, & Reilly 1992; Elliott 1992). Personality can also be represented in terms of a particular combination of underlying psychological traits, such as introvertedness/extrovertedness, shyness/assertiveness and the like, that influence the agents' behaviors under several circumstances (Rousseau & Hayes-Roth 1996); or it can be defined as a pattern of attribute-value pairs in which each attribute such as intelligence, amiability, strenght etc. is assigned a different numerical value for each agent (Goldberg 1997). As far as social behaviors are concerned, other characteristics have also been analysed and used for representing personality, such as competence, quirks, relationships, and so on (Reilly 1996).

We believe that for realizing more complex personalities, and personalities that are not necessarily related to typical emotional reactions or traits, it is essential to consider them as clusters of stable goals, and we want to explore the feasibility of this approach. In fact, personality is mainly considered by psychologists to be a coherent pattern of kinds of behaviors and interactions with the environment (Page 1983; Pervin 1989). This can be explained by assuming that the subject is pursuing some stable goals with different priorities, and possibly related to one another, which cause coherent patterns of behaviors across multiple contexts.

To face this problem, we propose a model of personality based on goals and plans. As explained above, personality is mainly concerned with the motivations of behavior; therefore we first want to find, for each possible personality we are interested in, the set of main goals that constitute it and regulate the agent's behavior. We also take into account the preferences that agents with specific personalities can have for different plans for achieving their goals.

In the following, goals are considered in the first place. We start from an AI work (Carbonell 1980) which has proposed to model human personality by using goal trees, and a psychological work (Ford 1992) that has provided a taxonomy of basic human goals.

Secondly, plans are analyzed: we propose that each personality has preferences for the use of some plans over others.

Then, the relevant features of our model of personality are summarized and some of its current shortcomings are outlined. Finally our model is instantiated into a set of personalities which influence a specific social behavior (help-giving).

Goals

Our model is inspired by (Carbonell 1980). As Carbonell notices, a personality could be defined as the set of actions characteristic and non-characteristic of it; but such actions are very many and can vary from one context to another. Therefore it would be impossible to list them exhaustively; furthermore, such a list would ignore the hierarchical relations among behaviors, and would not grasp its underlying motivations. In summary, a representation of personalities based on the set of typical behaviors would be too expensive and would lack generality. But since each action is based on underlying personal motivations, Carbonell proposes to use a more compact representation of personality traits under the form of a goal tree, where each goal would underlie general types of action. A "standard" tree is used for a prototypical person, where each goal is assigned a fixed priority with respect to others; each personality modifies such tree by changing the priorities assigned to one or more goals.

While we share the idea of using a goal structure for representing personalities, we do not need to represent the hierarchy for each personality as a "deviation" from the standard tree of a prototypical agent. In fact, Carbonell's work is related to the work by (Schank & Abelson 1977), who have drawn the attention on the need for representing plans, goals, and "themes" for a better comprehension of human behavior in story understanding. While in story understanding a standard tree is essential for making default inferences when no information about the character's personality is available, in designing BAs, we want that each agent is characterized by a marked personality, and therefore we do not mind about the relative importance of goals for a hypothetical standard agent. As a consequence, a different cluster of prioritized goals has been built for each of the personalities that have been chosen.

The set of goals from which to design those of our agents has been drawn from (Ford 1992; Ford & Nichols 1987). They have developed a psychologically-based taxonomy, with no hierarchical arrangement, where each category represents classes of goals at an abstract and decontextualized level of analysis. For example, the goal of entertainment, i.e. the goal of experiencing excitement and avoiding boredom, subsumes the goals pursued by behaviors like going to parties, seeking risk, avoiding daily routines, etc.

From that taxonomy we have chosen the following goals relevant for our purposes. Each description is taken from (Ford 1992) (Table 4.2, p. 88), except for the goal of Image (implied in another, wider category in Ford's taxonomy):

Resource Provision Giving approval, support, assistance, advice, or validation to others. Avoiding selfish or uncaring behavior.

Material gain Increasing the amount of money or tangible goods one has. Avoiding the loss of money or material possessions.

Social responsibility Keeping interpersonal commitments, meeting social role obligations, and conforming to social and moral rules. Avoiding social

transgressions and unethical or illegal conduct.

Belongingness Building or maintaining attachments, friendships, intimacy, or a sense of community. Avoiding feelings of social isolation or separateness.

Image Receiving positive evaluations from others, both in the realm of competence –being evaluated as skilful, intelligent, resourceful, etc.– and in the moral sphere –being evaluated as honest, generous, considerate, etc. Avoiding social disapproval.

Entertainment Experiencing excitement or heightened arousal. Avoiding boredom or stressful inactivity.

It can be empirically found that almost all of the listed goals have different priorities from one individual to another[2]; such goal priorities are a central feature of human personalities (Emmons 1989; Winell 1987). As already explained, from a BAs point of view, we need to design the clusters of prioritized goals for our agents.

Plans

There are of course many different ways (i.e. plans) for pursuing each goal: personalities manifest themselves not only by pursuing different goals, but also by assigning preferences to such plans. It is important to model such preferences because some goals may be common to several agents, but for maintaining believability it is better that the differences among their behaviors always emerge.

Our agents have a set of actions at their disposal which differ from one another not only for the main goals they achieve, but also for their side-effects, i.e. changes in the world that affect goals that are not actively pursued when the plan is executed. If the affected goals are important in the personality goal cluster, problems or opportunities may arise. For instance, the goal of "material gain", when instantiated in getting an interesting object from another agent, can be achieved in different ways (each of them is a plan involving the performance of several related actions): by stealing the object from the agent; by bargaining the object; by asking it to the agent. Each plan (when successful) produces the effect of having the desired object, but it differs from the others in terms of its side-effects: for instance the first plan has also the effect of violating social rules, thereby thwarting a goal belonging to the "social responsibility" category. Therefore, the "steal" plan would more probably be chosen by a selfish agent rather than an honest one, because the selfish does not care about the goal of respecting social rules.

[2]An important exception is f.i. the goal of "safety", which is expected to have the highest priority for everybody.

Relevant features and shortcomings of the model

In summary, in our model personality can be basically defined as:

- a cluster of goals with different priorities,
- a set of preferences over plans for achieving the goals.

It is important to note that goals may be pursued both actively and in reaction to suitable environmental contexts. In the latter perspective any occasion may be an opportunity for pursuing personality goals. As Sloman explains, "character and personality include long-term attitudes", which "are expressed in tendencies to make certain choices *when the opportunity arises*" (Sloman 1987) (original italics, p. 229).

Of course, many other aspects implied in the treatment of personality are not considered here. Among them are the criteria for goal activation and achievement, concerning the processes which regulate the pursuit of goals on the basis of evaluations about their relevance, importance and attainability (Ford 1992). Another missing feature concerns the so-called "goal-orientations" (Ford 1992), such as "maintenance-change", or "approach-avoidance", which refer to general personality-related styles of goal pursuit. Anyway, we believe that the features we have outlined are already a good starting point for representing and realizing complex personalities of BAs.

Personalities of Help-Giving

It would be impossible to define goal clusters for generic personalities, suitable to any situation or environment. Therefore, in order to instantiate our goal-based model of personality, we need to choose a situation or behavior in which to test several specific personalities. As a "case study" for realizing our agents, we have chosen the help-giving behavior, for several reasons.

First, help-giving is a very interesting complex social behavior. It is at the basis of exchange and cooperation, and involves the problem of "goal adoption" typical of multi-agent systems and of the works (either computational or not) on dialogue (cfr. the problem of perlocutionary effects, the so-called "master-slave assumption" underlying human-computer natural language dialogues, and the "benevolence assumption" discussed in Distributed AI). Secondly, help-giving implies sophisticated abilities of reasoning about others which are interesting by themselves (e.g., a help-giver could wonder about the reasons for the recipient's need, and decide not to help when the recipient is held responsible for it). Finally, the study of help-giving and of the behavior symmetrical to it (help-seeking) in multi-agent environments has already been pursued in some works of ours (Cesta & Miceli 1993; Rizzo, Miceli, & Cesta 1995), and their basic motivations and planning strategies have been explored.

Some personalities influencing help-giving

As for the agents' personalities, in previous works on help-seeking (Cesta & Miceli 1993) some "characters" have been outlined; in short, they consist in different modalities of help-seeking activated by particular goals. As for help-giving, some possible personalities influencing the modalities of help-giving are the following:

Altruist It sincerely cares about others, and is willing to help them, even at its own disadvantage.

Normative It is willing to help others if it should do it according to some norm (by role, reciprocity, equity), or at least if its helping behavior does not imply breaking any norm.

Selfish It cares only about itself; therefore it gives help to others only if it can profit from such a behavior (for instance, by offering its help in exchange for some reward).

Spiteful It enjoys making fun of others, by interfering with their plans, refusing to help, or playing tricks on them.

Goals influencing help-giving

Since believability is positively related with strong personalities, we have looked at the goals which can characterize strongly marked behaviors (of course, it has to be noticed that human personalities are less "marked" than those of our characters). The listed goals and their relative priorities are described in the following, with a more important goal preceded by a "+" and a less important one by a "–"[3]:

- Altruist
  - + Resource Provision. The altruist wants to give others what they need.
  - + Belongingness. This goal is useful to characterize the benevolence and kindness which usually accompany an altruist's behavior.
  - – Material Gain. The low importance attributed to this goal characterizes the marked behavior of a "strong" altruist.
  - + Image. The altruist likes to receive appreciation of its altruistic behavior; so, it welcomes others' positive evaluations, with special reference to the moral domain (being evaluated as generous, considerate of others, etc.)
- Normative
  - + Social Responsibility. The normative agent is very much concerned about social rules, and wants its own and others' behavior to comply with them.
  - + Image. The normative likes to receive appreciation of its normative behavior; so, it welcomes others' positive evaluations, with special reference to the moral domain. In addition, such a positive appraisal of a "normative image" can be viewed as a means for socially reinforcing norm compliance itself.
- Selfish
  - + Material Gain. The selfish wants primarily to acquire and maintain possessions.
  - – Resource Provision. The selfish is not interested in helping others, unless it is a means for its own "material gain".
  - – Social Responsibility. The selfish is ready to break rules in order to pursue the goal of material gain.
  - + Image. The selfish is interested in receiving others' positive evaluations, as a means for obtaining their adoption of its goals, with a consequent increase of its power (for instance, by presenting an image of "competence", it can be viewed by others as a good partner for exchanges); for the same reason, the selfish might be interested in hiding its selfishness, to avoid negative evaluations in the moral domain (for instance, being viewed as a potential cheater).
- Spiteful
  - + Entertainment. The spiteful likes to play jokes at others; it finds amusing that others get into trouble.
  - – Social Responsibility. The spiteful does not care very much about social rules.
  - – Resource Provision. Since it takes pleasure in others' troubles, the spiteful is unlikely to help others.
  - – Image. The spiteful does not care about its social image and is likely to risk negative evaluations in view of making fun of others and interfering with their plans.

To give a more concrete idea of the behaviors typical of the listed personalities, and to explain how the listed goals can influence such behaviors, here are some examples of possible interactions between BAs endowed with the given personalities, and a human agent (HA) that asks them to help her wash the dishes in an every-day life scenario. The examples described below are based on communicative actions, but many other non-communicative behaviors could be typical of such agents:

Altruist : (going toward the kitchen) "Don't worry! I will do all the dishes by myself". The Altruist is always eager to help, given the high priority assigned to the goal of resource provision.

Normative : (simply standing still) "Today it's your turn!". The Normative wants that the high important goal of social responsibility is respected by HA.

[3] At the moment, a goal is given as generically more (or less) important than others, without specifying to what extent).

Selfish : (looking at HA's pocket) "I might help you with the dishes if you give me one dollar". The Selfish may accept to help if it is in exchange for some reward, in order to pursue its high importance goal of material gain.

Spiteful : (smiling) "I'll do them before mom arrives..." The Spiteful knowingly deceives HA and gives no help, in order to make her get into trouble, and hence take pleasure in this.

Implementation through deliberative and reactive planning

The ideas described above are being verified in an implementation of agents that interact in a text-based modality with a human user in a virtual environment. The goal of the project is to produce "believable" help-giving behaviors according to the introduced personality traits.

The implementation is being accomplished within Prodigy (Veloso *et al.* 1995), a means-ends, state-based, backward-chaining planner, and RAP (Firby 1987), a reactive execution system that executes the plans in adaptation to the changeable features of the environment.

Different problem solving operators are represented in terms of personalities of the agents. Prodigy can narrow the search space in several different ways, namely through the use of functions to constrain value selection for operator variables and by using domain-dependent control rules that are applied to constrain the decision choices in the planning algorithm. A control rule is an if-then rule that specifies which choices should be selected or avoided given the current state and meta-level planning information. "Select" and "reject" control rules can be used to prune the search space, while "prefer" ones can determine the order of exploration of the choices. By using constraint functions or applying control rules to the choice of operators or to the bindings for the parameters of operators, Prodigy is able to build different plans for each of the different personalities by taking into account their different goals and preferences over operators and bindings.

The plans produced by Prodigy are given as input to RAP, which executes them according to the priorities assigned to the plans on the basis of the differences among personalities.

In Firby's system, a RAP (Reactive Action Package) is the basic execution unit. It associates one or more program-like sets of actions (i.e. uninstantiated plans, called methods), that can be executed for performing a particular task (i.e. achieving a goal), with the situations in which they are applicable. In other words, each task is mapped on a particular RAP, which specifies a set of methods for achieving it and their conditions of applicability. The initial tasks to be performed (which must be set by the designer or taken from a sketchy plan formed by a deliberative planner) are put on a task agenda. The interpreter selects one task at a time from the agenda by considering task selection constraints (e.g. explicit ordering or temporal constraints) and heuristics (e.g. higher priorities or closer deadlines). If the selected task has not yet been achieved, the RAP associated with it is taken from the RAPs library, and an applicable method is chosen for execution. If the method is a primitive action, it is executed; otherwise all of its subtasks are put on the agenda, and the main task is put on the agenda to be executed again after all of its subtasks are finished.

In our RAP-based implementation, the goals listed in the previous section are represented as tasks in the agenda with different priorities for each personality. The plans for achieving the tasks are represented as methods available in the RAPs associated with them.

The variability among agents' behaviors due to their personalities is therefore produced in two ways: first, by the Prodigy's selection of operators according to the interaction between their side-effects and the goals which are important for each personality; second, by the RAP interpreter's activation of different tasks depending on their priorities (e.g. the "entertainment" task can be pursued by any agent at any time, but it will be pursued more often by the Spiteful agent).

Conclusions

In this work our goal-based model of personality has been outlined, and we have described the specific domain of our investigation concerning the realization of believable help-giving behaviors, the way the goal-based model may be used in that domain, and the implementation grounded on the Prodigy planner and the RAP reactive execution system.

The use of a deliberative planner for producing personality-based plans, and the integration of the planner with a reactive system for plan execution, allow to flexibly represent the different helping personalities of our social agents. Future developments of this work include a tighter integration of planning and execution, which would allow to model changes in the personalities of the agents according to the feedback coming from the kinds of interactions occurring between them and human agents.

References

Bates, J.; Loyall, A. B.; and Reilly, W. S. 1992. Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.*

Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* July.

Carbonell, J. 1980. Towards a process model of human personality traits. *Artificial Intelligence* 15:49–74.

Cesta, A., and Miceli, M. 1993. In search of help: Strategic social knowledge and plans. In *Proceedings of the 12th International Workshop on Distributed Artificial Intelligence.*

Elliott, C. 1992. *The Affective Reasoner: A process model of emotions in a multi-agent system.* Ph.D. Dissertation, Northwestern University. Technical Report no. 32.

Emmons, R. A. 1989. The personal striving approach to personality. In Pervin, L. A., ed., *Goal Concepts in Personality and Social Psychology.* Hillsdale (NJ): Lawrence Erlbaum Associates. 87–126.

Firby, R. J. 1987. An investigation into reactive planning in complex domains. In *Proceedings of the 6th National Conference on Artificial Intelligence.*

Ford, M. E., and Nichols, C. W. 1987. A taxonomy of human goals and some possible application. In *Humans as Self-Constructing Living Systems: Putting the Framework to Work.* Hillsdale (NJ): Lawrence Erlbaum Associates. 289–311.

Ford, M. E. 1992. *Motivating Humans. Goals, Emotions, and Personal Agency Beliefs.* Newbury Park (CA): SAGE.

Goldberg, A. 1997. Improv: A system for real-time animation of behavior-based interactive synhtetic actors. In Trappl, R., and Petta, P., eds., *Creating Personalities for Synthetic Actors.* Berlin: Springer.

Hayes-Roth, B.; Brownston, L.; and vanGent, R. 1995. Multi-agent collaboration in directed improvisation. In *First International Conference on Multi-Agent Systems.* Cambridge (Mass.): MIT Press.

Laurel, B. 1991. *Computers as Theatre.* Reading (Mass.): Addison-Wesley.

Maes, P. 1995. Artificial life meets entertainment: Lifelike autonomous agents. *Communications of the ACM* 38(11).

Page, M. M., ed. 1983. *Personality. Current Theory and Research.* Lincoln: University of Nebraska Press.

Pervin, L. A. 1989. *Goal Concepts in Personality and Social Psychology.* Hillsdale (NJ): Lawrence Erlbaum Associates.

Reilly, W. S. 1996. *Believable Social and Emotional Agents.* Ph.D. Dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. Technical Report CMU-CS-96-138.

Rizzo, P.; Miceli, M.; and Cesta, A. 1995. On helping behavior in cooperative environments. In *Proceedings of the International Workshop on the Design of Cooperative Systems (COOP'95)*, 96–108. Le Chesnay Cedex (France): INRIA.

Rousseau, D., and Hayes-Roth, B. 1996. Personality in synthetic agents. Technical report, Knowledge Systems Laboratory, Department of Computer Science, Stanford University, Stanford (CA).

Schank, R. C., and Abelson, R. P. 1977. *Scripts, Plans, Goals and Understanding. An Inquiry into Human Knowledge Structures.* Hillsdale (NJ): Lawrence Erlbaum Associates.

Sloman, A. 1987. Motives, mechanism, and emotions. *Cognition and Emotion* 1(3):217–233.

Veloso, M.; Carbonell, J.; Perez, A.; Borrajo, D.; Fink, E.; and Blythe, J. 1995. Integrating planning and learning: the prodigy architecture. *Journal of Experimental and Theoretical Artificial Intelligence* 7:81–120.

Winell, M. 1987. Personal goals: The key to self-direction in adulthood. In *Humans as Self-Constructing Living Systems: Putting the Framework to Work.* Hillsdale (NJ): Lawrence Erlbaum Associates. 261–287.