A Cognitive Model for the Generation and Explanation of Behavior in Virtual Training[1]

Maaike Harbers [ab] Karel van den Bosch [b] Frank Dignum [a]
John-Jules Meyer [a]

[a] *Utrecht University, P.O.Box 80.089, 3508 TB Utrecht, The Netherlands*
[b] *TNO Defence, Security & Safety, Kampweg 5, 3796 DE Soesterberg, The Netherlands*

1 Introduction

Virtual systems have become common training instruments in organizations such as the army, navy and fire services. In most of the current systems instructors play a major role; they enter the effects of trainee actions into the system, play the role of other characters in the scenario, and provide instruction, guidance and feedback to the trainee. Cognitive models can be used to support instructors by autonomously generating virtual character behavior. However, without knowing the reasons for an action, it is more difficult for a trainee to understand the situation and learn from it. Some virtual training systems with explanation components have been proposed, e.g. Debrief [3] and the XAI system [5, 1]. These components explain a character's actions by giving information about its state at the time it performed an action. However, the explanations do not provide the character's underlying motivations for its actions.

We propose a cognitive model based on BDI theory (belief desire intention) that is able to generate character behavior and explanations. Because the character reasons and plans with BDI concepts, the model is also able to generate explanations in terms of the characters's beliefs and goals. Such explanations in high level concepts are easy to understand for humans. Moreover, they not only provide information about the character's state, but yield insight in the reasons why it performed a certain action.

2 The Cognitive Model

Our approach resembles planning methods based on hierarchical task networks (HTNs) [4]. In HTN planning, an initial plan describing the problem is a high-level description of what is to be done. Plans are refined by action decompositions, which reduces a high-level action to a set of lower-level actions. Actions are decomposed until only primitive actions remain in the plan. In our model, initial plans, action decompositions and primitive actions correspond to main goals, divisions of goals into sub-goals and actions, respectively. Figure 1 represents the cognitive model of a simple virtual fire-fighter.

The virtual characters in a training scenario fulfil specific roles with corresponding goals and tasks. For example, a dispatch center operator should properly inform others, a policeman ensure order and safety, and a fire-fighter bring incidents to a successful conclusion. The characters maintain their overall goal during the complete training session. To achieve their main goal, they have to adopt proper sub-goals, sub-sub-goals, and finally perform the corresponding actions. The selection of goals depends on a character's beliefs and the relation between a goal and its sub-goals. Relations differ for example on the amount of sub-goals to be achieved in order to achieve a goal (one or all), or the order in which sub-goals have to be achieved (fixed or free order). The fire-fighter in our example has the beliefs **Fire** and **Victim**. Because it also has the belief

[1]The full version of this paper will appear in the proceedings of ExaCt 2008, ECAI Workshop on Explanation Aware Computing. This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

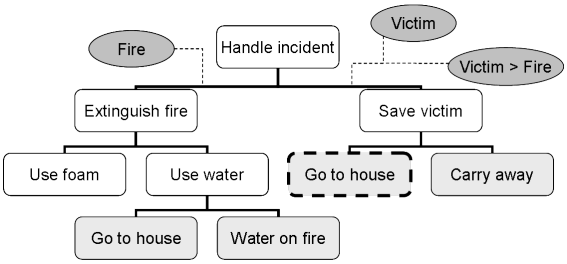

Figure 1: The cognitive model of a fire-fighter

that extinguishing a fire has priority over saving a victim (**Victim** $>$ **Fire**), it adopts the goal **Save victim** and executes the action **Go to house**.

For the generation of behavior, the cognitive model is used from top to bottom and the result is a sequence of actions. For the generation of explanations, the model is used from actions to the underlying sub- and main goals. In an explanation, the beliefs and goals that were responsible for a particular action are made explicit. However, a long trace of beliefs and goals may underly one action, but a trainee often does not need all information. A selection mechanism is needed to filter the most useful information out of the many goals and beliefs. A possible heuristic is that an action which is part of different goals in the task hierarchy is explained by the beliefs that were responsible for the selection of the current goal and not the other. For example, the action **Go to house** in Figure 1 can be part of two goals and is explained by the character's beliefs **Victim** and **Victim** $>$ **Fire**.

The implementation of the cognitive model requires an agent programming language in which beliefs and goals are explicitly represented. Moreover, because explanations often involve goals *and* beliefs, it should be possible to combine the two and reason with them. In the current BDI-based agent programming languages, agents cannot reason about their own goals and beliefs. This has been avoided because modifications to beliefs and goals in an agent's reasoning process might result into undesired loops. We have chosen to implement our model in 2APL [2] and make updates of the agent's goals and beliefs in its belief base. The updated goals and beliefs are strictly separated from the 'normal' beliefs and do not influence the generation of actions. They are only used for the generation of explanations.

3 Conclusion

We have proposed a cognitive model for the generation and explanation of behavior. The use of explicit goals and beliefs in an agent's reasoning process distinguishes our model from most other approaches of behavior generation. Moreover, explanations generated by other accounts of explaining agents do not refer to the agent's beliefs and goals.

References

[1] M.G. Core, T. Traum, H.C. Lane, W. Swartout, J. Gratch, and M. van Lent. Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82(11), 2006.

[2] M. Dastani. 2APL: a practical agent programming language. *Autonomous Agents and Multi-agent Systems*, 16(3):214–248, 2008.

[3] W. Lewis Johnson. Agents that learn to explain themselves. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 1257–1263, 1994.

[4] S. Russell and P. Norvig. *Artificial Intelligence A Modern Approach*. Pearson Education, Inc., New Jersey, USA, second edition, 2003.

[5] M. Van Lent, W. Fisher, and M. Mancuso. An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of IAAA 2004*, Menlo Park, CA, 2004. AAAI Press.