Evaluating the Believability of Virtual Agents with Anticipatory Abilities

(Extended Abstract)

Quentin Reynaud[1,2,3]
[1] Sorbonne Universités
UPMC Univ. Paris 6
UMR 7606 - LIP6, 4 Place Jussieu,
F-75005 Paris
quentin.reynaud@lip6.fr

Jean-Yves Donnart[2]
[2] Thales Training and Simulation
1 rue du général de Gaulle
95523 Cergy Pontoise, France
jean-yves.donnart@thalesgroup.com

Vincent Corruble[1,3]
[3] CNRS, UMR 7606
Laboratoire d'Informatique de Paris 6
4 Place Jussieu, F-75005 Paris
vincent.corruble@lip6.fr

Abstract
In this paper we present an experimental evaluation of a generic model of cognitive process allowing virtual agents to exhibit anticipatory abilities. This model is included in an agent architecture for agent-based simulations of virtual cities. With user experiments, we show that this mechanism brings about an improvement in the efficiency of the behavior, and check that external observers are able to perceive it. We also confirm that this improvement in efficiency leads, up to a point, to an improvement in believability as judged by human observers. Beyond this level of efficiency, believability reaches a plateau and can even decrease.

Categories and Subject Descriptors
I.2.11 [Distributed Artificial Intelligence]: Intelligent agents.

General Terms
Design, Experimentation, Human Factors, Theory.

Keywords
Virtual agents, agent reasoning, cognitive module, anticipation, behavior believability, user evaluation, agent-based simulation.

1. INTRODUCTION
In this paper we tackle the issue of enriching the decision process of virtual agents with anticipatory abilities. Our claim is that these abilities increase the behaviors efficiency of the virtual agents, and consequently the believability of these behaviors. We also claim that agents which are perceived as too efficient (just as agents not efficient enough) could be perceived as unbelievable. We support those claims by integrating the corresponding module (which is presented in [1] and takes some inspiration from [3]) in a flexible hybrid agent architecture [2] and by evaluating it, with a subjective user-based experimentation focusing on the perceived efficiency and believability of the agent's behavior.

Appears in: Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.

2. EXPERIMENTATIONS

2.1 Protocol
The experiments were run with the simulator described in [2]. All simulations took place in the same virtual environment, "The Republic Square" in Paris, France, which represents an area of 1.6km². About 20 actions are feasible, such as "eat in a restaurant", "drink coffee", "sit on a bench", etc. The user experiments were conducted through an online survey. These experiments involved 144 participants.

2.2 Hypothesis
The goal of these experiments is to validate the following hypothesis. H_1: "the presence of the anticipatory module improves the behaviors efficiency as perceived by the participants". H_2: "the presence of the anticipatory module improves the behaviors believability as judged by the participants". H_3: "an improvement in the behaviors perceived efficiency brings an improvement in the behaviors perceived believability".

2.3 Behavior efficiency and believability
The participants had to compare two short video sequences (around 1 minute) both showing the proceedings of an agent's morning. When the simulation starts, the agent is at home (at 8.30 am), then it goes to work and stays there until noon. The video also highlighted the current time of the simulation, the current action of the agent, its current satisfaction levels (satisfied, slightly unsatisfied, unsatisfied, strongly unsatisfied), and its schedule (working hours). The only difference between the 2 videos is that, in the second one, the agent does have an anticipatory module, while the agent in the first video does not.

	Average score	Standard deviation	Mode
Without anticipation	**3.69**	1.60	4
With anticipation	**4.64**	1.74	6

Table 1: efficiency results from behaviors with and without anticipation, on a Likert scale from 1 (lowest) to 7 (highest).

The first goal was to evaluate the perceived efficiency of the behaviors shown in both videos on a Likert scale. Table 1 contains a summary of the 144 collected answers. These results show a

clear gain in efficiency: the average score of efficiency is around 1 point better with anticipation than without, and the mode is increased by 2 points. A Student's t-test confirms that this gain is significant (with a confidence level below 0.01). These results confirm H_1. The second question was about the believability assigned by the participants to the same behaviors.

Table 2 shows these results:

	Average score	Standard deviation	Mode
Without anticipation	**4.44**	1.53	5
With anticipation	**4.96**	1.65	6

Table 2: believability results from behaviors with and without anticipation, on a Likert scale from 1 to 7.

These results are not as clear-cut, but do not leave doubts concerning the impact of anticipation about believability. The average believability score is around 0.5 point better with anticipation, and the mode is increased by 1. A Student's t-test also confirms the significance of these results (confidence level below 0.01), which validate H_2. According to the comments left by the participants, the main reasons for these increases in efficiency and believability are based on two major differences between the behaviors of the two agents. First, the anticipatory agent is able to leave home early and reach its work place on time. Secondly, although it is just a little hungry when it leaves home for work, the anticipatory agent is able to anticipate the fact that it will not be able to eat during its working hours, and that at noon it will be very hungry, and uncomfortable with it. So, it decides to eat something on the way to work, when it walks by a cafe.

2.4 Behavior believability as a function of behavior efficiency

We decided to consider the evolution of the behavior believability as a function of the behavior efficiency (figure 1), in order to verify H_3. To do so, we gather the results of the two previous questions, which produce two sets of 144 couples of efficiency and believability scores. From these 288 couples, we calculate the average believability score associated with each grade of efficiency. In other words, we collect all the couples with an efficiency score equal to 1, and calculate the average score of believability of these couples. And we iterate the process for the 6 other values of efficiency. It produces 7 average believability scores, one for each efficiency grade. This diagram can be divided into two zones. In the first one, with an efficiency ranging from 1 to 4 (low and medium efficiency), each gain in efficiency brings a proportional gain in believability. Above an efficiency score of 4, little or no improvement is obtained, and one could hypothesize the beginning of a decrease in believability. This means that, when a behavior is considered as highly efficient, an additional gain in efficiency does not bring an additional gain in believability, and can even impact it negatively. These results support H_3, but only partially. To conclude, one can say that H_3 is confirmed, but only when the perceived efficiency is low. Beyond that point the relation between efficiency and believability is unclear. One could propose the hypothesis that, in a manner analogous to the "Uncanny Valley", we might observe what one might call a "Canny Hill" in the observation of behavior believability as a function of its efficiency.

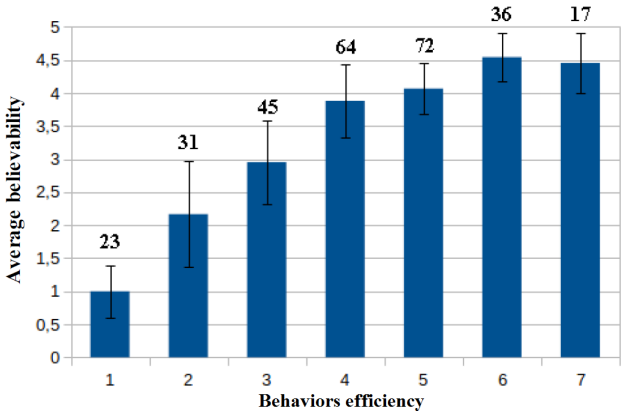

Figure 1: average behavior believability as a function of the behavior efficiency, with the standard deviation (vertical segments), and the number of results (numbers at the top of each column)

3. CONCLUSION

In this paper we presented some user experiments focused on the measure of perceived behaviors efficiency and believability. We confirmed that the anticipatory module brings an improvement in the behaviors perceived efficiency and verify that external observers find these behaviors more believable than pure reactive ones. Furthermore, we showed that there is a link between the efficiency and the believability of a behavior, and that when a behavior is considered as moderately efficient, increasing its efficiency brings a gain in believability. When the behavior becomes more efficient, this link is not verified anymore. In conclusion, we can argue that adding anticipatory abilities is a crucial step toward increasing agents' believability even if we highlighted some limitations to this result. There are various issues yet to be addressed. For example, one could compare the behaviors of our agents with the behaviors of a human-controlled avatar, thanks to another user-based experiment. That way, we would be able to see how far from real human behaviors we are in terms of believability. One could also propose some more complex experiments, by increasing the complexity of environment, actions, and agents.

4. REFERENCES

[1] Quentin Reynaud, Etienne De Sevin, Jean-Yves Donnart and Vincent Corruble. A cognitive module in a decision-making architecture for agents in urban simulations. In AAMAS 2012 Workshop on Cognitive Agents in Virtual Environments (CAVE), Valencia, Spain, 2012.

[2] Etienne de Sevin, Quentin Reynaud and Vincent Corruble. FlexMex: Flexible Multi-Expert Meta-Architecture for Virtual Agents. In Advances in Cognitive Systems, Palo Alto, USA, 2012.

[3] A. Doniec, R. Mandiau, S. Piechowiak, and Espié S, Anticipation based on constraint processing in a multi-agent context. Autonomous Agents and Multi-Agent Systems, vol. 17, no. 2, pp. 339–361, 2008.