

Intelligent Agents Living in Social Virtual Environments – Bringing Max Into Second Life

Erik Weitnauer, Nick M. Thomas,
Felix Rabe, and Stefan Kopp

Artificial Intelligence Group, Bielefeld University, Germany
{`ewitnau, nthomas, frabe, skopp`}@`techfak.uni-bielefeld.de`

Abstract. When developing cognitive agents capable of interacting with humans, it is often challenging to provide a suitable environment in which agent and user are really co-situated. This paper presents a straightforward approach to use Second Life as a persistent, “near natural“, and socially rich environment for research on autonomous agents in complex surroundings, learning social skills, and how they are perceived by humans. We describe an integration of our conversational agent Max, and his underlying cognitive architecture, with the Second Life environment. As a result, Max is able to autonomously navigate and multimodally communicate with other avatars inside the Second Life world. Observations from a first trial and reactions of human residents to the virtual agent are presented.

Key words: social virtual environments, intelligent agents, second life

1 Introduction and Background

In 1992, Neal Stephenson introduced in the novel *Snowcrash* [1] the terminus *Metaverse* to describe a social virtual environment, where humans interact as avatars in a 3D world that uses the metaphor of the real world. Today, there are several incarnations of such Metaverses [2–6] already populated by millions of human users. While not the oldest but certainly among the most well-established, *Second Life* is a large “3-D virtual world created by its users” as Linden Lab, the company running the technical infrastructure of Second Life, states. At the time of writing these lines, there are more than 13 million *residents* (registered users of Second Life) of which more than 430.000 have been online during the last week [7]. Every day, Second Life gains new residents who are spending time and money to socialize, shop, explore and maintain friendships. People are even found immersed in this world to an extent that they are marrying, renting flats, or building their own houses there, and they show communication behaviors that resemble those known from the real world [8].

While already being employed for novel multimedia and e-learning applications [15, 14], a persistent social environment of the vast dimensions as the Second Life world also offers many new possibilities for A.I. research. First of all, it is rich enough with respect to opportunities for exploration and interaction,

the level of detail of the underlying world simulation, as well as the dynamics of change in the world. At the same time, it is intuitive and easy to enter and use for both humans *and* for intelligent agents. Crucially, it provides a place of social encounter for humans and intelligent agents in which both meet at eye level as they join each other in the form of avatars - either controlled by a human resident or by the agent’s mind. An agent can thus easily reach the same basic perceiving and acting capabilities as the human residents, which enables to overcome many of the limitations and unbalances of other interactive settings. For example, in human-robot communication a social interaction between man and machine is often hindered by shortcomings of the robot with respect to elemental perceptive abilities. On the other hand, many A.I. scenarios employ sparse virtual worlds to co-situate humans and intelligent agents, which faces problems of acceptance and generalizability. The Second Life world, however, provides a high degree of richness and simulation fidelity while proving ease of implementing, e.g., processing speech input or 3D vision. The former benefits from the fact that the majority of conversations is done using the text chat, the latter from direct access to 3D world object-data. The fact that all objects and avatars within Second Life possess a unique id also simplifies the identification of objects and avatars.

In this paper we present work on bringing Max, an intelligent virtual agent developed at Bielefeld University, into Second Life (SL, henceforth). We show how the agent’s underlying cognitive architecture is connected to the SL infrastructure via a mediation layer, and we present observations from Max’s first interactions with SL residents.

2 Bringing Intelligent Agents Into Second Life

2.1 Prerequisites

There are a number of general prerequisites an intelligent agent must meet if it is to autonomously join, act, and communicate in a complex and dynamic social environment. To name the most important ones, the agent has to

Perceive. Basic perception must include the possibility to locate and identify structures and avatars. Recognition of attributes like appearance or movement of objects should be possible, just like the perception of speech (and preferably gestures and facial expressions) in interaction with other avatars. Also, a minimum of self-perception like connection status and the state of his body representation in the virtual world is helpful.

Act. The agent must be able to navigate and move around in the virtual world. It must have the means of expressing itself by speech, gestures and facial expressions. Further, the agent should be able to manipulate the world in a physical sense.

Reason. The agent must build beliefs about the world, it should have goals and inference mechanisms. An extensible world knowledge is important in a dynamic environment with interactions governed at large by common-sense conven-

tions. Additionally, in persistent environments, a memory is required for keeping track of events over time to enable learning.

Social Skills. Built upon the previously mentioned abilities, the agent must possess basic social skills. This pertains to knowledge and actions pertaining to social conventions and protocols, e.g., navigating while respecting social distance, turning and gazing to establish joint reference, or maintaining group closure. Ideally, the agent would be able to recognize and reason about other’s beliefs and intentions. Further, the agent should have emotions [10] and, at best, personality traits to appear as a sociable interaction partner (e.g., cf. [16]).

We argue that SL is a well-suited environment for investigating and attaining these skills in intelligent agents. The fact that SL is a rich social world in that human-human interaction is easily found and technically observable, and human-agent interaction is easy to engage in, can foster approaches to learning and testing these skills in actual interactions. We also stress that our focus is not just on how to serve and take advantage of the existing SL user community, but also on how to use the SL technology independently of the official server grid as an advanced and controllable setting for intelligent agent research. In fact, there are implementations of the SL server readily available [11] that can be set up for own purposes.

2.2 The Conversational Agent Max

The Artificial Intelligence Group at Bielefeld University has been developing the virtual human Max [13, 17] to study how natural conversational behavior of humans can be modeled and made available for A.I. systems. Aimed at mixed-initiative dialog and collaborative interaction in dynamic virtual environments, Max rests upon a general cognitive architecture that combines abilities for concurrent perception, rational reasoning and deliberation, emotion, and action.

Perception and action are directly connected through a reactive component, affording reflexes and immediate responses to situation events or input by an interaction partner. Input is gathered via an arbitrary number of sensors, which feed into a perception component that utilizes sensory buffers and provides elementary sensory data integration. Reactive processing is realized by a behavior generation component in charge of realizing all behaviors that are requested by other components. This includes feedback-driven reactive behaviors like gaze tracking the current interlocutor and secondary behaviors like eye blink and breathing. Additionally, this component in the normal system provides real-time realization of multimodal utterances into synthetic speech and synchronized animations of the agent’s face or body.

The deliberative processing component determines when and how the agent acts, either driven by internal goals and intentions or in response to incoming events which, in turn, may originate either externally (user input, visual perception) or internally (changing emotions, goal achievement, belief assertion). All deliberative processes are carried out by a BDI interpreter [12], which continually pursues multiple, possibly nested plans (intentions) to achieve goals (desires) in the context of up-to-date knowledge about the world (beliefs).

An emotion system continuously runs a dynamic simulation to model the agent’s emotional state. It receives input from both the perception (e.g., seeing a person triggers a positive stimulus) and the agent’s deliberative component. The calculated emotional state is available anytime, e.g., as a weighted emotion category that is mapped to Max’s facial expression and is available for the agent’s further deliberations.

The version of Max we have started from here is the one employed in a public computer museum [13]. There, Max tries to engage visitors in conversations in which he provides them with information about various topics of interest. Max takes as input German natural language typed in on a keyboard, and he responds with a synthetic German voice and nonverbal behaviors like manual gestures, facial expressions, or gaze. That way, he performs mixed-initiative dialogs in which he is capable of initiating, holding, resuming and releasing topics and dialogue goals. In doing so, Max is also capable of small talk and responds to questions, interruptions, or topic shifts.

All of the interactive and conversational capabilities of Max are modeled within the BDI framework. Drawing on dedicated plans, desires, and beliefs, this component interprets all incoming events, decides how to react dependant on current context, and produces appropriate responses. At present, the system comprises about 2.000 plans, most of which implement condition-action rules that test either user input or the content of a dynamic knowledge base; their actions can alter the dynamic knowledge structures, raise internal goals and thus invoke corresponding plans, or trigger the generation of actions. In result, Max’s deliberative component combines pattern matching techniques to model robust small talk about large domains with consistent plan pursuit to conduct longer, coherent activities and to act proactively. Since all plans are formulated over general beliefs, desires, and actions, this approach naturally lends itself to integration of conversational skills with other behavior plans, as needed for an agent that is to act and interact autonomously in the SL environment.

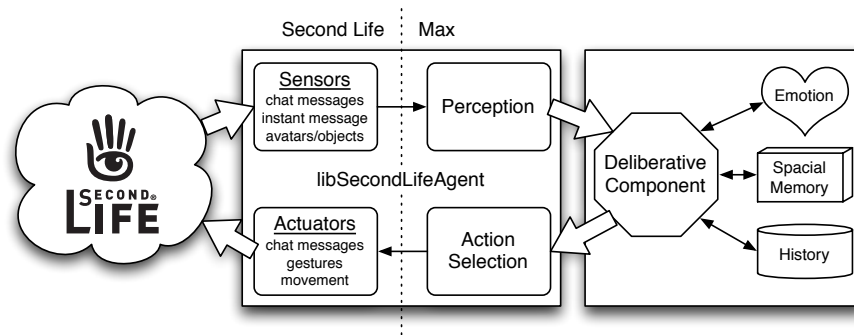


Fig. 1. A detailed view on the software architecture: The libSecondLifeAgent acts as a Second Life client for Max.

2.3 Connecting Max With Second Life

The two main challenges in bringing Max into SL were, first, extending Max's deliberation system with behavior plans adequate for the new environment and, second, providing a mediation layer for communication with the Second Life servers. For the latter we chose the open source library *libsecondlife* over the official Linden Lab client as a basis for our implementation, since it already provides a considerable set of functionality.

The connection of Max's architecture with SL is realized via a dedicated module (*libSecondLifeAgent*) as shown in figure 1. Building upon *libsecondlife*, simulated sensors receive data from the server and feed it into Max's perception component. This includes object and avatar positions, status updates, avatar appearances, avatar profiles, chat and instant messaging, as well as changes of friends and inventory status. The deliberative component, running the BDI interpreter, has been extended to keep track of the objects and avatars surrounding Max. In pursuing intentions, the deliberative component can invoke actions specified in terms of a text to utter, a conversational function of the overall dialog act (used to pick appropriate nonverbal behaviors), and, now, additional functional descriptors of avatar actions. An action selection component in the SL client maps these specifications onto the actions provided by the SL environment. Possible actions encompass movements like jumping, walking, sitting down, dancing, flying, and teleporting. The scope of communication-related actions include chat and instant messaging, gesturing, facial expressions, and gazing at objects or avatars. Actions relevant for social life include management of appearance, monetary resources, the inventory and friends list.

Upon these extensions, Max was able to enter the environment and communicate multimodally with human avatars and objects. He has access to over 100 in-built gestures (including facial expressions), autonomously navigates through the virtual world, follows avatars, and can access all user functions available in the official SL client. Figure 2 shows two screenshots of Max (or, more correct, his avatar) in SL.



Fig. 2. Left picture: Max's avatar (left) and a human's avatar (right) engaged in a conversation. Right: A typical situation of Max meeting a group of Second Life residents socializing.

3 Max’s First Steps Into Second Life – Observations

Max’s first steps into SL were mainly conducted in two of the biggest German communities in SL, the *Apfeland* and the *Germania* regions. To test the feasibility as well as the acceptance of our approach to embed a virtual interactive agent in the SL social environment, we have entered SL with Max on a Friday and Sunday evening, which are times with an above-average number of residents logged-in. As Max in his current version lacks the ability for social navigation, SL residents who were by no means aware of Max’s artificial nature, were specified as a “conversational target” and then autonomously approached and engaged in a conversation by Max. To that end, Max retrieves the position of the nearby resident’s avatar, approaches her, initiates a chat, and tries to sustain the conversation as long as possible. During the conversation, Max constantly checks the distance between his avatar and the avatar he is chatting with. Max takes a conversation to be finished if the other avatar walks away or expresses her intention to end the dialog. After the conversation, we carefully briefed the resident and asked how the conversation was experienced. All of Max’s first steps were automatically logged to text-files as well as captured to video, for later analysis and improvement.

Observations of Max’s interactions with residents show that human users, although assuming Max’s avatar to be controlled by a human, are not always willing to engage in a conversation. In total, Max approached 15 residents out of which 5 were responded while the others were already engaged in other activities. The conversations Max autonomously had with residents lasted from 114 to over 600 seconds, with an average of around 300 seconds, and turn-switches ranging from 8 up to over 50 (average 25). In the post-briefing, residents reported that they suspected the agent to be controlled by an artificial system at a certain point during the conversation, or they got frustrated without realizing that they were not interacting with a human and thus ended the conversation. The most common reasons residents gave were repetitive agent behavior, too general or ambiguous answers, or too quick responses.

A comparison with experiences from user interactions with Max in the museum setting [13], in which the visitors are aware of interacting with a machine upfront, shows behavior similar to what we see arising in SL residents after or (possibly) during their uncovering of the artificial nature of their interaction partner. Notably, testing and fastening down their interactant’s capabilities. One remarkable concern articulated by a majority of residents was that a mixed society, in which human and machine-controlled avatars are so undistinguishable as in SL, would lead to distrust and disharmony within the community.

4 Conclusion

In this paper we have presented work toward utilizing existing social virtual environment technology for research on intelligent virtual agents. While being meant as a proof of concept, by connecting Max’s cognitive architecture to the

Second Life server, we have tapped new resources for research on interactive and autonomous machines. Constructing social virtual worlds like SL, we can now simulate scenarios that have a similar degree of complexity and dynamics as the physical world, but where real humans and agents perceive and act on an equal level. This is a prerequisite to our future research on social interaction of humans and artificial entities in mixed virtual societies. For one thing, SL offers great opportunities for learning-based approaches to modeling social skills, both by interacting with humans or other agents, as well as by observing other avatars interact. Our first observations are promising in that they show that intelligent agent techniques can be employed with and benefit from these settings, but also do they raise a number of new research issues, e.g., regarding the acceptance of machines that are potentially indistinguishable from humans.

References

1. N. Stephenson: Snowcrash. Bantam Books (1992)
2. Second Life, <http://secondlife.com/whatis/>
3. Active Worlds, <http://www.activeworlds.com>
4. Forterra Inc., <http://www.forterrainc.com>
5. There, <http://www.there.com>
6. Multiverse, <http://www.multiverse.net>
7. Second Life Economic Statistics, http://secondlife.com/whatis/economy_stats.php
8. Friedman, D., Steed, A., Slater, M.: Spatial Social Behavior in Second Life. In: Pelachaud, C., Martin, J.-C., André, E., Gérard, C., Kostas, K., Pelé, D. (eds.) IVA. LNCS, vol. 4722, pp. 252-263. Springer, Berlin (2007)
9. libSecondLife, <http://www.libsecondlife.org/wiki/MainPage>
10. Becker, C., Kopp, S., Wachsmuth, I.: Why emotions should be integrated into conversational agents. In: Nishida, T. (eds.) Conversational Informatics: An Engineering Approach, chap. 3, pp. 49-68. Wiley (2007)
11. OpenSimulator For Second Life, <http://opensimulator.org/>
12. Huber, M.J.: JAM: A BDI-Theoretic Mobile Agent Architecture. In: Proc. Autonomous Agents'99, Seattle (1999)
13. Kopp, S., Gesellensetter, L., Krämer, N., Wachsmuth, I.: A conversational agent as museum guide—design and evaluation of a real-world application. In: Intelligent Virtual Agents. LNAI, vol. 3661, pp. 319-343. Springer, Berlin (2005)
14. Hayes, E. R.: Situated Learning in Virtual Worlds: The Learning Ecology of Second Life. In: AERC Conference Proceedings (2006)
15. Eliëns, A., Feldberg, F., Konijn, E., Compter, E.: VU @ Second Life – creating a (virtual) community of learners. In: Proc. EUROMEDIA 2007, L. Rothkrantz and Ch. van der Mast (eds.), Eurosis, pp. 45-52 (2007)
16. Breazeal, C.: Towards sociable robots. Robotics and Autonomous Systems 42 (3-4): 167-175 (2003)
17. Lessmann, N., Kopp, S., Wachsmuth, I.: Situated Interaction with a Virtual Human - Perception, Action, and Cognition. In: Rickheit, G., Wachsmuth, I. (eds.) Situated Communication, pp. 287-323, Mouton de Gruyter (2006)