Investigating the role of memory in interactions between intelligent virtual characters and humans over time

Karla Bransky and Debbie Richards
Computing Department
Macquarie University
North Ryde, NSW, 2120
Australia,
richards@science.mq.edu.au

ABSTRACT
Memory is an important ability in aiding human interactions. It allows us to build interpersonal relationships by helping the development of trust and sharing of knowledge. Our ability to remember names, faces, places, events and other details enriches our ability to communicate. It therefore seems intuitive that any intelligent virtual character interacting with humans over a period of time should exhibit a similar ability for recall. Memory in virtual characters has typically been implemented as a solution to virtual characters remembering information considered essential for their success in their intended roles. This is frequently achieved by predefining the information that agents retain by making assumptions about the kind of information that agents should either remember or forget. However, human memory is not restricted to purely pertinent information. We are thus investigating the role of memory over time by conducting experiments to determine what sort of information, type of memory (e.g. episodic and/or semantic) and level of accuracy are considered most appropriate by humans to be recalled by intelligent virtual agents with whom they are interacting in order to find a property for lease or sale. Results should give us a better idea of how to increase the believability of virtual characters using long-term memory as well as greater understanding of the expectations that humans have of agents.

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

General Terms
Measurement, Design, Experimentation, Human Factors,

Keywords
embodied conversational agents, intelligent virtual agent, long term memory, longitudinal studies.

1. INTRODUCTION
Memory is an important ability in aiding human interactions. It allows us to build interpersonal relationships by helping the development of trust and sharing of knowledge. Our ability to remember names, faces, places, events and other details enriches our ability to communicate. It therefore seems intuitive that any intelligent virtual character interacting with humans over a period of time should exhibit a similar ability for recall.

Memory in virtual characters has typically been implemented as a solution for virtual characters to remember information considered essential for their success in their intended roles. This is often achieved by predefining the information that agents retain by making assumptions about the kind of information that agents should either remember or forget. However, human memory is not restricted to purely pertinent information. While memory, both short and long term, are basic components of computer-based systems, we note that storage of predefined data may be what the system developer implements but this may not meet the expectations of the users who find themselves having to repeat information they have already provided in previous conversations. Such frustration and lack of apparent sensitivity on the part of the computer-based system will not endear the agent to the human and encourage ongoing interaction. Furthermore, people continuously build on their mental model of the world as it is impossible to know what kind of information will be needed in the future. Restricting information to a limited subset or domain diminishes the ability to build long-term relationships, and thus the believability of a virtual character.

Therefore, a project has been formulated with the aim of gaining a better understanding of the kinds of things that users expect virtual characters to remember, and how users react to agents that exhibit memory from contexts wider than their design role. The investigation will be achieved by implementing various levels of recall ability in a virtual real-estate agent, similar to that of Cassell and Bickmore’s [12, 13] REA. The real-estate domain has been chosen as people usually have multiple interactions with a real-estate agent over a period of time in their search for a property, and successful real-estate agents are able to remember their clientele from interaction-to-interaction. Specifically, they would remember details directly relating to the sale (like requirements, preferred locations, budget etc.) and also those details that seem less significant, but are important for maintaining rapport with the client. The various models of memory implemented in the agent will be tested in a series of controlled experiments. Users will interact with the agent over several weeks, with the agent exhibiting different levels of recall ability. We expect that models recalling the greatest amount of information will be judged as the most believable. We are interested in how the agent’s recall of irrelevant information contributes to the interaction, as well as user’s reactions to correctly or incorrectly recalled information.

In summary, we are investigating the role of memory in interactions with humans and intelligent virtual characters over time. Results should give us a better idea of how to increase the believability of virtual characters using long-term memory as well as greater understanding of the expectations that humans have of agents. In the next section we consider the background and relevant literature. In section 3 we discuss the types of human memory relevant to our project. In section 4 we present our approach and the domain in which our studies will be conducted. In section 4 our approach and experiments are described. Conclusions appear in section 5.
2. BACKGROUND

Intelligent Virtual Agents (IVAs) have been the focus of much research within the field of Artificial Intelligence in the past 20 years [18]. An IVA is a piece of software, generally considered to be autonomous in some way, which imitates the behavior of a human or animal and is embodied within a virtual environment [9]. In the tradition of Turing [28], a primary aim in the field of Virtual Agents is the creation of believable characters that are useful in their situated paradigm. Agents that are believable are generally considered to be those that will be the most effective. Believable agents are not necessarily realistic, but rather, behave in a manner that appears convincing and natural to the user, whilst serving their intended purpose. An example of such an agent is Clippy; Microsoft’s much hated office assistant. Clippy’s default appearance was a paperclip with Groucho Marx style eyebrows. Clippy was by no means realistic, but he was certainly believable. Only with the support of other agents did this become true [29]. Believable agents are useful in their situated paradigm. Agents that are believable are more likely to be used as this is what we hypothesize users will expect agents to exhibit [26].

The believability of an IVA can be affected by many aspects. It can be enhanced by behaviors such as breathing, gaze, gesture, speech, emotion, social skills and intelligence. Or, it can be adversely affected by factors such as poor implementation or easily recognizable behavior patterns [17].

An important aspect for human interaction is our ability to store, retain, recall and organize information. This ability is known as memory and it is essential to our basic functioning as human beings. For instance our ability to remember time and events allows us to keep track of what we have done, and to make plans for the future. Many architectures, both agent and cognitive, include memory modules but not much research has been done specifically in establishing how memory contributes to interactions between users and IVAs.

Memory in virtual agents has typically been implemented to address the issue of how agents remember information from interaction to interaction. This is considered necessary for the agent to effectively carry out the role for which they are designed, within their various paradigms (games, narratives, assistive computing, educational agents, etc.). For example, a pedagogical agent needs to remember past lessons held with a student in order to function appropriately in their role [10, 18]. Likewise, an office assistant like Clippy should be able to recall that a particular user already knows how to write a letter, and prefers to seek help when they need it [26]. For this reason it is clear that agents that interact with users over a period of time need the ability to remember interactions with users, and depending on their role, the ability to recall and act on knowledge gained from those interactions.

From the literature, it is not clear how much information IVAs should remember. It is typically assumed that agents should only store information that is specific to their purpose. However, this is problematic. For example consider the following questions; how do you draw the line at the level of detail agents should be able to recall? How do we determine the potentially useful information from the irrelevant? Should irrelevant information be discarded? It is important to note that human memory is not restricted to storing only task-oriented information. Humans continuously update their mental model of the world and even the most seemingly insignificant detail may come in handy in some future situation. A simple solution to this problem then would be to store all the information available. However, this is equally problematic and not just because of obvious issues of limitations on the size of the storage space. In an agent that remembers everything there are ethical issues to consider; particularly those of data protection and privacy [29]. Vargas et al [29] propose that any memory model for long-term interactions should incorporate ethical methods of regulating the kind of information that is encoded, remembered and “forgotten”. They are currently working on a socially-aware memory for companion agents and have developed an initial prototype memory built on top of FaTiMA [20]. They plan to test it for the kind of information that should or should not be disclosed. Additionally, over time information decreases in value, becoming outdated and irrelevant. Forgetting is an essential coping mechanism in human memory which allows information that has not been retrieved in while to gradually be forgotten. Although work has been done in implementing forgetting into memory architecture in IVAs, it has not yet been explored whether agents should exhibit a similar kind of knowledge loss. Human memory is by no means perfect, and is subject to influence and error. This is a phenomenon which we as humans are explicitly aware of and we show understanding towards others when memory lapses occur. However, will users expect IVAs to exhibit perfect recall due to their computational basis, or will they attribute the same ‘flawed’ memory processes to IVAs as they do other people?

Another question of interest is how do users respond to agents that exhibit false memories? Should agents present memories of their own as if they were real humans? It is commonly thought that users will feel cheated or deceived by agents with personal human histories [7]. However preliminary work on engagement and deceit by Bickmore, Schulman and Yin [7] found that agents that tell autobiographical stories from the first person perspective were found to be more engaging, and more likely to encourage the user to continue to use the system. This is an indication that false memories can increase the believability of an IVA. Therefore, with these questions in mind, we propose to conduct a study that investigates the role that memory plays in IVAs which interact with users over time.

3. HUMAN MEMORY

Human memory is not a single unitary system, but an array of interacting systems which perform several important functions particularly; learning, organizing and remembering, forgetting, repression, storage and retrieval. In the psychological study of memory there is considerable agreement that it can be roughly divided into three general types, commonly known as the Atkinson-Shiffrin model [1]: sensory memory, short-term memory (STM) and long-term memory (LTM) [4]. LTM can be divided into the further distinctions of episodic and semantic memory [27]. Episodic memory is our memory for events and stories like “Last week I went for a job interview” and “I went to my cousin’s wedding in June” whereas semantic memory is our memory for general knowledge about the world like the meaning of words, names of capital cities and where the local cinema is located. Usually when people talk about “memory” they are referring to episodic and semantic memory; the recollections of what they have learned and their experiences [22]. For this reason we are focusing our study on both episodic and semantic memory, as this is what we hypothesize users will expect agents to exhibit over long-term interactions.
4. REAL ESTATE DOMAIN & APPROACH

The domain of a real estate agent was chosen by Cassell et al [14] because of the amount of talk with social function it produces. Particularly as “in real estate sales, a good agent will continue to focus on building rapport throughout the relationship with a buyer” (Garos, 2000 in [5]). The choice of the domain of real estate agent for our project echoes this sentiment, but in particular, that due to the broad nature of conversation customers are likely to have with a real estate agent means that is likely that the agent may be able to gain information not directly pertaining to the predominant communicative goal. However, our second reason outweighs the first, in that due to nature of the domain, the relationship between clientele and the agent extends past the first interaction over a longer period of time. Therefore, in order to build rapport and maintain relationships with their clients, real estate agents need to recall details from past interactions with their clientele.

There are two kinds of information that a real estate agent particularly needs to recall. These are task-based details that directly relate to the sale (such as the number of rooms required, whether clients have kids, what location they prefer, how much they are willing to spend etc.) and social-based information essential for building rapport. Social-based information includes facts such as the client’s favorite colour is blue, that they hate dogs, and that they talked about how they spent time at the beach on the weekend. This kind of information is important not only for conversational reasons; but also for the agent to build up a better mental model of the client as well as establishing trust. Thus, what would be stored for each user would differ as different matters would arise in each situation which could not be anticipated. It is clear that in this case it would be difficult to predefine all the information that may arise in an interaction, and thus it is a good domain for the investigation.

The agent will be modified to exhibit recall and forgetting of task-related and social-based information including both semantic and episodic memory. This will then be tested in a series of experiments with users over a period of time, where the agent will exhibit different levels of recall ability. It is expected that the scenarios where the agent exhibits recall for the greatest amount of information will be judged as the most believable. We are interested in how the agent’s recall of irrelevant information contributes to the interaction, and secondly, how such information can be used to update the agents model of the world. Due to its particular relevance to the domain and their use of conversational agents we next present REA in more detail.

4.1 Overview of REA

Rea is a virtual real estate agent developed by the Gesture and Narrative Language Group at MIT Media Lab. Rea was developed as a platform for conducting research into embodied conversational agents [5]. Rea is an embodied, real-time conversational interface agent which uses conversational protocols in an effort to make interactions as natural as if face to face with another person. Users are sensed passively through cameras and audio input, and communicated with by Rea producing speech combined with intonation, facial expressions and gestural behavior. Rea is displayed on a large projection screen and user’s head and hand movements are tracked using cameras. Audio input is captured by microphone worn by the user. Rea is shown below in Figure 1.

Although initially Rea was only capable of Eliza-like responses [39] which echoed back the user’s input, Rea was continuously improved until about 2005. In her latter stages Rea was able to interact autonomously as a real-estate agent: she could answer questions about properties contained in a database and gave users a virtual tour of the house. To achieve this Bickmore and Cassell [5] constructed a discourse planner which allowed small talk and task talk to be interleaved during the initial interview. This planner used an activation network-based approach and the architecture afforded the smooth transition from deliberate (planned) behavior to opportunistic (reactive) behavior. The architecture allowed the pursuit of multiple goals and each node in the network represented a conversational move that Rea could adopt.

Rea responds to user’s verbal and non-verbal input and can conduct a conversation describing the features of each property. She allows users to interrupt her by recognizing when they make cues usually associated with conversational turn-taking, and then continues at the next available point in the conversation. When Rea misunderstands what a user says she able to initiate error correction and displays the respective nonverbal output (such as facial expressions, gestures and voice output). Her responses are generated by a natural language generation engine which has been modified to include synchronized gestures. The speech acts the users are engaging in are determined by a simple discourse model. During task talk, the questions that Rea might ask about are the users buying preferences (e.g. number of rooms needed) and during small talk she might talk about the weather, tell stories about herself, the lab or real estate or about other events and objects in her shared physical context with the user. Conversational moves are planned to reduce the face threat to the user and maximize trust, whilst pursuing Rea’s goals as effectively as possible. However, Rea’s responses are still quite limited; in small talk modes much of the content is ignored and dialogue moves forward purely because the user has responded. Whenever the closeness with the user needs to be increased, Rea chooses to do small talk and many task-related topics are governed by preconditions that mean they cannot be broached until a certain activation level has been reached.

Figure 1 (Figure 1: REA, the virtual Real Estate Agent, Cassell & Miller, 2007)
4.2 Memory in Rea
Rea appears to have been designed for once-off conversations. No attempts have been made to implement any kind of user recognition, or long term memory in the agent. Semantic knowledge about the world in pre-encoded, determined by what the programmers thought it was likely that any person living in the area would know. Neither does it appear that any attempts were made to clarify whether the information that was implemented in the agent seemed natural to the users. Likewise Rea cannot extend her knowledge of the world past the short-term knowledge that she gains via her conversational interactions. Any knowledge about the client gained is lost at the conclusion of the conversation and would not be recalled should the same user interact with Rea in the future.

Rea’s communicative content is drawn from the structure of the context, private and shared knowledge, syntactic knowledge and the lexicon. The background knowledge base is organised with the sources of information that Rea and the user share i.e. their common ground and also describes the relationship between Rea’s private information and the questions that are of interest to the user. Semantic and Syntactic constraints of utterances are also contained in the KB. Throughout the conversation dynamic updates from the discourse model are sent to SPUD so that the current state and context is kept up to date. Updates include the shared common ground and the current attentional state of the conversation. Using the communicative goal, KB and the updated conversational context, utterances are built by SPUD element by element. Once this is complete the information is then sent to the Action scheduler for execution and the behavior is then exhibited by Rea.

5. AIMS & METHODS

We are undertaking a study investigating the role that memory plays in interactions with humans over periods of time. To do this, we are implementing a virtual real estate agent, similar to Cassell & Bickmore’s REA. This agent will be used in a series of experiments with users over a period of time (most likely no longer than four weeks) in which various levels of ‘memory’ will be implemented into the agent. Our primary aim in this study is to formulate a clearer idea of what kinds of things users expect agents to remember, and how users react to agents that exhibit memory from a domain wider than their design role.

5.1 Creating the agent for the experiment

We intend to develop the virtual real-estate agent by modifying a suitable previously developed agent framework. Due to the domain overlap, REA might appear to be a good first choice. However, as explained above, REA is a complex communicative system including speech recognition, vision, task-oriented and social dialogue, as well as multimodal natural language generation. Even if we had access to REA, for the purposes of our studies we do not wish to complicate or delay the experiment by setting up an experimental environment which involves input processing and multimodal interpretation and a broad range of potential side effects in human interpretation of the ECA stemming from parameters such as the ECA’s voice, facial expression, body language etc.

Bickmore, Shulman and Shaw [8] make the point that it is hard to make progress on a particular research question, when each researcher/project needs to build all the supporting infrastructure involved in a fully functional embodied conversational agent. In response to this problem they offer a lightweight agent using DTalk and the DLite body. Currently, our intention is to avoid reinventing the wheel and to use functionality in the Virtual Human Toolkit\(^1\), namely the SmartBody module to develop our embodied conversational agent.

We have also considered the use of ION and FAItMA [2, 3] as this framework will allow the incorporation of emotional appraisal into the storage and retrieval of memories and related work has been done incorporating autobiographical memories into Fatima (e.g. [2], [16], [21]). However, our initial project is only short-term (1 year) and implementing our character using Fatima does not seem necessary or best use of our constrained time. We do however intend to suggest how Fatima might be used or extended, if necessary, to incorporate our findings.

As a final alternative, in previous work we have developed an agent-based virtual environment known as BOSS (BOrder Security System) [25] which involves an airport world and related characters which could be adapted for the real estate domain. We would prefer, however, to increase our contribution to the community by employing and giving feedback on the (re)usability of other ECA technology.

5.2 Experimental requirements and setup

At this stage, our requirements are that an ECA will be displayed on a computer screen and deliver a number of sequences of dialogues to the participant. At certain intervals the participant will be asked to respond. The information entered will be used to create the dialogues to be presented in subsequent sessions. What information is (mis)used will depend on the experimental condition. Interaction during the session will be via mouse and keyboard. The environment will be web-based. Making the agent accessible online ensures that accessing the agent is more convenient for the participants, and thus allows both a greater participation rate as well as reducing the rate of participant dropout. This is particularly important in studies that involve multiple interactions over time (Bickmore & Schulman, 2009). The agent will be modeled on behavior typically exhibited by real-estate agents. For example the agent should exhibit behavior such as:

- asking clients about what their requirements are in a property,
- suggesting properties, and
- sales-based behaviors like being friendly, professional, etc.

5.3 Experimental design

We will use the agent in a series of experiments over the period of a few weeks, with participants interacting with the agent multiple times. Participants will be asked to play the role of a prospective buyer or renter of a property. Undergraduate students are the likely participants to be recruited due to their accessibility. Participant experience with buying and renting will be one of the factors considered in the analysis of the data.

Participants will be asked to evaluate the interactions they had with the agent using a likert scale to indicate their level of satisfaction with 1) the agent’s behaviour; 2) the agent’s

\(^1\) http://vhtoolkit.ict.usc.edu/index.php/Main_Page
appearance; 3) the agent’s believability; 4) the agent’s ability to remember relevant facts; 5) the overall experience/interaction. Users will be asked to specifically answer 1) if the agent appeared to remember them; 2) if so, what did they do or say that gave this impression; 3) what aspects of the dialogue they found irrelevant or annoying; 4) what other information they believe the agent should have remembered. User evaluation and experimental data will be gathered primarily via questionnaires, although we may also conduct some observation of interaction (perhaps via a webcam). By providing all studies online, data collection is also more convenient.

The study will explore at minimum two conditions: a control condition in which the agent remembers nothing besides the property selection criteria specified by the client, and the experimental condition in which the agent recalls at least one piece of (additional) information from a previous interaction, and exhibits that recall to the user. It may be necessary to differentiate between relevant and irrelevant additional material and thus the following experimental conditions are desirable:

- The agent recalls at least one relevant piece of information from a past interaction with that user
- The agent recalls at least one irrelevant piece of information from a past interaction with that user, as well as relevant information
- The agent recalls at least one irrelevant piece of information from a past interaction with that user and no relevant information

Initially we will focus on the simpler goal of comparing no additional information with additional information. We intend to use a 'Repeated Measures' design with two within subjects factors (Scenario, NoRecall/Recall) and one between subjects factor (stimuli order). In a 'Within Subjects Design' each participant receives all conditions instead of only, in contrast to a 'Between Subjects Design'. A 'Within Subjects Design' is considered to be more powerful [23] in regards to conclusions that are drawn from associated data analysis if responses of participants under different conditions are positively correlated. Generally speaking, the variance of difference scores tend to be less in a 'Within Subjects Design' than the variance of the observations on which they are based [11]. Furthermore, in a 'Within Subjects Design', data for two (or more) treatments (stimuli) is gathered from the same participant rather than comparing the performance of one or more participants under different conditions with each other, as is the case in a 'Between Subjects Design' where each participant only receives treatment with one stimulus. Moreover, a 'Within Subjects Design' reduces errors that are due to individual differences between participants by comparing the performance of the same participant(s) under the influence of all possible stimuli.

In the present study this means that all participants will receive two similar scenarios/dialogues in the same experimental session. One scenario will not include any additional information and the other scenario will include comments and suggestions by the IVA containing evidence of having recalled what the participant had said in a previous interaction. The participant will rank the two scenarios along a number of criteria such as which interaction they found more enjoyable, more natural or more useful. They will be asked to explain their responses and explicitly ask if they found any of the conversation irrelevant, annoying, enjoyable, etc.

As part of each experimental session and within the context of each scenario, further casual conversation will occur to be used in later experimental sessions.

As a pre-experimental step, our agent will need to “get to know” the participant. For this purpose we will ask the participant to enter their personal details and the type of property they are looking for but also include a number of “small talk” questions like “What did you do over the weekend?” or “How is the weather outside?” The user will be able to answer these questions as free text and we would be looking for more detailed responses than REA would be expecting. In the case of REA, as in many socially-aware agent systems, the goal is to appear socially-capable, so that the user will feel more positive towards the agent and the interaction. For example, in learning environments using pedagogic agents it has been observed that the learner’s perception of the learning experience is positively affected when a lifelike character is included in a computer-based interactive learning environment. This has become known as the persona effect (Lester et al. 1997).

We anticipate a total of 4 sessions, including the getting to know you session, per participant with orderings of conditions changed in each alternate session and new scenario/dialogues each time.

6 CONCLUSION

We have described a project which seeks to gain a better understanding of the role that memory plays in interactions between users and virtual agents. We expect the project will show that a better implemented memory model increases the believability of IVAs. Using data from a number of experiments to be conducted with participants over a one month or longer period, we will be able to gauge if the participants find the IVA more believable and appealing over time if they demonstrate better recall.

Many studies have been conducted which look at the importance of various factors in creating believable and acceptable agents for purposes such as teaching, training, guidance, trouble-shooting and support. While some (e.g. [6]), consider whether humans will be willing to interact with the agents over time, most do not. It is possible that favourable responses after one interaction may not persist when technology novelty has worn off or the obvious lack of real human-intelligence and social ability is observed.

In this study we look at the long-term use of agents and how the recall of information supplied by the participant not pertinent to the original interaction may make the agent seem more believable and improve the experience and maybe even the outcome. It is interesting to note that agents can assist human memory retention and knowledge transfer, as found by Moreno et al. [24] when the learner was assisted by a pedagogic agent compared to a computer-based text environment without the agent. It will be interesting to see if such behaviors in the agent, i.e. memory retention and knowledge transfer, also benefit human learners and human-agent interaction in general.

7. REFERENCES


