Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back

Cyril Brom
Charles University in Prague
Dept. of Software and Computer Science Education
Malostranske nam. 25
Prague 1 – 118 00 – CZ
brom@ksvi.mff.cuni.cz

Jiří Lukavský
Academy of Sciences of the Czech Republic
Institute of Psychology
Politických vězňů 7
Prague 1 – 110 00 – CZ
lukavsky@praha.psu.cas.cz

ABSTRACT
Recently, it has been proposed that virtual characters should have a *full* episodic memory storing more or less everything happening around them, as opposed to an *ad hoc*, that is, *special purpose* episodic memory. However, it was not much clear, what exactly this "fullness" should mean. The purpose of this paper is to clarify it and show how it can contribute to the agents' believability. Later, our work-in-progress applying several aspects of the full episodic memory will be reviewed. At the time of writing, the memory model integrates following parts: a hierarchically organised memory for events, a component reconstructing the time when an event happened, a topographical memory, and an allocentric and egocentric representations of locations of objects. The main functional features include: representation of complex episodes (e.g. cooking a dinner) over long intervals (days) in large environments (house), forgetting based on emotional importance of episodes, and development of search strategies for objects in the environment.

Categories and Subject Descriptors
I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents. I.2.6 [**Learning**]: Connectionism and neural nets, Knowledge acquisition.

General Terms
Algorithms, Design, Experimentation, Theory.

Keywords
Virtual characters, episodic memory, autobiographic memory, spatial memory, dating of events, allocentric and egocentric representations.

1. INTRODUCTION
A *believable virtual agent* is an autonomous agent who seems lifelike, whose actions make sense to the audience, and who allows them to suspend their disbelief providing convincing portrayal of the personality they expect or come to expect (Loyall, 1997). It contributes highly to the believability of an agent if the audience is able to establish *empathic relations* with the agent (e.g. Paiva et al., 2004). In other words, the users should be able to spontaneously and naturally tune themselves into the agent's "thoughts" and "feelings" (Baron-Cohen, 2003, p. 21), to perceive that the agent is experiencing or about to experience emotion (Paiva et al., 2004). Arguably, episodic memory is one of the key components contributing to establishing the empathic relations, because it allows the user to understand better the agent's history, personality, and internal state: both actual state and past state. It has been already discussed that believable agents (or characters) should have, at least for some applications, episodic memory (Ho & Watson, 2006; Castellano et al., 2008). In our previous work, we have even proposed that they should have a *full* episodic memory (Brom et al., 2007). But what does it mean "a full episodic memory" (FEM)? In the above mentioned paper, we used a vague definition of a memory storing almost everything happening in the proximity of the agent, as opposed to the *ad hoc/special purpose* solutions. Certainly, this full episodic memory *cannot* be a faithful reconstruction of human episodic memory—it can be a model mimicking some of its features, but which ones? And when speaking about empathic characters, are there some features that are more important for them than others?

The main purpose of this paper is to revisit the notion of the FEM, give it a more exact shape and reconcile it in the light of needs of empathic agents. The aim is to arrive a) at a tentative list of features of episodic memory most important for empathic agents, and b) at the definition of the FEM.

We begin our search tapping at the door of people who should be most knowledgeable about real episodic memory: psychologists and neurobiologists. It will turn out, however, that we won't be much lucky. Then, we will sketch out some cognitive skills requiring some aspects of episodic memory. This step will help us with the objective (a), but only partly with (b). Through another step, we will come very close to the definition of the FEM, but, surprisingly, we will resist the temptation to define it claiming that the definition would be of no use. But we will also arrive at a definition of something else, more important than the FEM.

After this discussion, the paper will give a technical context to some of the ideas sketched out previously reviewing briefly our on-going work on a virtual character that encodes and recalls complex events, including detail information about time and space. An important feature of our model is a gradual forgetting. For the space constraints, the model cannot be detailed here fully, but the reader can find more in (Brom & Lukavský, 2009). The

Cite as: Title, Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

whole paper addresses primarily the audience of developers of empathic virtual characters; it aims at providing them with some hints concerning equipping their agents with episodic memories. However, some points may be of use also for neuro-/psychologists. The discussion will be kept on the conceptual and the methodological levels. This paper extends our original work on characters with the FEM (Brom et al., 2007) and complements our methodological paper on possible utilisation of virtual characters with episodic memory in the field of neuro-/psychological computational modelling (Brom & Lukavský, 2008). The conceptual issues related to virtual characters with episodic memory (not necessarily a full one) have been also discussed by Ho & Watson (2006).

2. TOWARDS FEATURES OF THE FEM

The important concept behind current neuro-/psychological memory research is the idea of multiple memory systems. Episodic memory (Tulving & Donaldson, 1972; Baddley et al., 2001) is an umbrella term for those of these systems that operate with representations of personal history of an entity, which entails encoding these representations, their maintenance, consolidation and recollection. These representations are related to particular places and moments, and connected to subjective feelings and current goals. Fundamentally, the episodic memory is being distinguished from the semantic memory and the procedural memory. The former is conceived, more or less, as systems operating with general facts about the world as viewed form the objective perspective. The latter covers processes related to skill learning and the subjective experience is again not emphasised. The importance of agent's subjective history makes episodic memory an interesting area for empathic agents developers.

However, beyond these general statements the issues become dim. For example, to which extent the systems of episodic, semantic and procedural memory overlap? Many accept the tentative neuro-/psychological taxonomy of memory types developed by Squire & Zola-Morgan (1991) (see also Eichenbaum & Cohen, 2001), but this taxonomy elaborates the notion of procedural rather than episodic memory. Some make a distinction between episodic memories consisting of sensory-perceptual-conceptual-affective information derived from single experiences, and autobiographical knowledge, which is basically personal semantic knowledge, devoid of context in which it was acquired (Conway, 2005; Williams et al., 2008). The terminology does not seem to be settled yet, therefore it is not possible to simply implement the properties of human episodic memory. Think of this example: If a virtual agent remembers that her glasses are at the TV, is this related to episodic memory (a remembrance of the episode of putting this glasses there), semantic memory (the general knowledge about where things tend to be), or procedural memory (an unconscious stimulus–response-like habit)? A neuropsychologist would likely say that all the three alternatives are possible. But which of these properties should FEM possess? It is not only a problem of psychological terminology. Imagine we know how to implement the agent with the ability to recall the position of glasses – what is actually recalled? Think of the first alternative. Should she recall only the relation `<at, glasses-23, TV-4>`? Or also the features of TV, for instance its colour? Should she also recall that she put the glasses at the TV *because* she wanted to read newspaper, a task she needed different glasses for? What should happen, if, after all, the glasses are not at the TV? To our knowledge psychological details of these processes sufficient for implementing our virtual agent are not available.

As we are speaking about the needs of believable characters, we can, for obvious methodological reasons, undergo the "user centric" turn and to stop asking questions about the nature of episodic memory and to start asking questions about what users would expect from FEM agents. Assuming they would expect from them the same as from real humans, we are actually asking questions about users' folk psychology. The problem is, that at least to our knowledge, it is not known much about this issue. Nevertheless, it seems reasonable to expect that most people do not have the concept of episodic memory at all and there are suggestions that humans expect the human memory in general to behave unlike it really behaves (e.g. Loftus, 1979; Friedman, 1993).

Hence, the neuro-/psychology thread helped us to reveal two problems with our hypothetical FEM: that 1) we do not know what features the FEM should possess, and 2) even if we knew it, we would not know how to implement them. It seems that we will have to guess the features and somehow try to implement them, a blind search approach. Luckily, even though neuro-/psychology cannot offer us the technical specification for the FEM, it can constrain our search. It can offer us some interesting general architectures (e.g. Conway, 2005; Zacks et al., 2007), inspiring observations, e.g. the idea of false memories (Loftus, 1979; Brainerd & Reyna, 2005), and some hints such as that one has to distinguish between a short-term and a long-term memory (that is, briefly, between memories from which information fades out quickly vs. not so quickly[1]). And of course, this discipline can offer us loads of data, from which are arguably most interesting for our purposes diary studies (e.g. Wagenaar, 1986; Burt et al., 2003), event perception studies (Shipley & Zacks, 2008) and forensic psychology data (e.g. Loftus, 1979). It offers us also some computational models of laboratory tasks such as memorising of words or navigation in the Morris water maze (e.g. Miyake and Shah, 1999; Norman et al., 2008; Krichmar et al., 2005), but we would hardly utilise these for the FEM, unless we aim at engaging our agents in really weird tasks. Finally, we know that we should evaluate our models on users, that is, we should ask whether the models would pass an episodic memory variant of the Turing test.

What next? Perhaps... could we try the luck at the very field of virtual agents? Indeed, several reports have emerged during last years on agents with various episodic memory-like capabilities. Agents have been reported with spatial memory to increase believability of navigation and/or "what-where" judgments (Thomas and Donikian, 2006; Strassner and Langer, 2005; Peters, 2006; Isla and Blumberg 2002; Noser et al., 1995). Other characters have been equipped with a memory for past events for the purposes of debriefing (Johnson, 1994; Rickel and Johnson, 1999; Dias et al., 2007). Also there has been work on robots with a simple episodic memory (Dodd, 2005) and work at the intersection of the field of virtual characters and the artificial life investigating

[1] What exactly means "quickly" depends on the kind of memory one is talking about. One story would be told by a neurobiologist investigating memory mechanisms at a neural level (e.g. Kandel, 2001), another by a psychologist investigating memory for words (Baddley, 1986; Chap. 3). One may also argue that humans do not have one short-term memory and one long-term memory, but many interacting memory systems, each of which keeps information over a specific time interval.

how different types of episodic memories can improve an agent's chances of survival (Ho et al., 2008).

These models depart from computational neuro-/psychological models in one important way. They are aimed at representing complex, rich, human-like episodes, or large spaces such as a city with many landmarks and objects. If a forgetting mechanism is implemented, the models can be used in scenarios lasting long time intervals, e.g. days. However, these models can not be conceived as FEM models; they are technical, special-purpose solutions invented to address a particular issue (and they typically work well for the purposes of that issue). Can they help us to underpin the features of the FEM at the least? Yes, similarly to the neuro-/psychology, we can draw inspiration from them; however, the standpoint is now different. These models force us to think not about the properties of the FEM, but about cognitive skills an agent potentially may have that demand these properties. In other words, we are forced to think about how to utilise the FEM.

2.1 How to utilise episodic memory?

On the one hand, we are still not far from where we begun, on the other hand, we have some vague ideas, hints and constraints, which encourage us to try the good-old-fashion approach: brute-force search. Let us now challenge the notion of FEM during a two-step search. First, we will ask "why": why we need an FEM agent? We will lay down a tentative list of cognitive skills that demand some kind of episodic memory, not necessarily the FEM, and ask for examples of real world applications that would utilise agents with particular skills (see? this step is motivated by the outcome of that part of our previous debate that concerned itself with virtual characters). Of course, applications featuring FEM agents have to demand *all* the skills, and we will try to identify these applications. Second, we will ask "what": what requirements on the FEM architecture stem from these skills (this will capitalise on neuro-/psychological inspirations).

Now, let us start with the "why" part—the required agent skills:

A1. Debriefing. Tutoring agents should be able to talk about history of given lessons. As said above, agents with this ability already appeared.

A2. Giving information. This skill extends A1 for the purposes of long-living agents; it is the ability of giving users information about what happened in the virtual world in the past. Arguably, this skill is presently most important for role-playing game (RPG) characters. Predominantly, these agents now tend to inform players about important past happenings by means of pre-scripted dialogs. It would be useful to generate this information dynamically both from the design point of view as well as for believability reasons. Virtual characters living in large yet-to-be-developed social virtual worlds (Goertzel, 2007) would need this ability as well.

A3. Remembering the course of interaction. Agents with conversational abilities, such as virtual companions (Castellano et al., 2008), virtual guides (Kopp et al., 2005; Lim, 2007) or again NPCs need to keep a track of the dialog with a user. Long-term companions may be engaged in dialogs extended over many days. This may demand building information about their users. Think of an agent chatting with an elderly user about her old photographs (Companions, 2006); the agent should remember when the events portrayed had happened and who they are about.[2] Note, that this ability is, to a large extent, based also on semantic memory system.

A4. Searching for objects. Think again about the example of searching for glasses. Every agent living in a world that include objects that can change their positions beyond the agents' capabilities must able to judge reliability of contradictory memory records (unless the agent looks directly to the world map). Where are the glasses: at the TV, or at the bed side table? Suppose the agent needs also a pencil, which may be either at the TV, or somewhere in the study room. Where the agent should go first? [3]

A5. Topological orientation. Agents embodied in virtual environments (as opposed to speaking heads etc.) should be able to orient themselves, no matter whether they act in a city, a family house, or a country-side. This is an easy issue. However, whenever the topology can change dynamically, the agents have to construct dynamically their internal "topological memories". Even though there is an abundance of work addressing this issue in robotics (e.g. Kuipers, 2000), and some also in the domain of virtual characters (e.g. Thomas and Donikian 2006), many players of real-time character-based strategies are still witnessing soldiers "hiding" behind a once-existing wall that has been destroyed, for the place was marked as a cover by a designer.

A6. Mental imagery and predictions.[4] Agents employing declarative representations may be already conceived as using imagery. One general example is the usage a graph of way-points during path-planning, but there are also many special purpose imagery-based tasks virtual agents may need to solve in specific applications. For example, some tutoring agents may be required to answer questions such as: "what would happen if I press this button?" While simple answers may be represented in advance, for more complex situations, the agent may need to generate the answer using the imagery (cf. Rickel and Johnson, 1999).

A7. Sharing of knowledge. It is known that sharing of information can improve agent's survival (e.g. Ho et al., 2008; Cace & Bryson, 2007). Generally, this objective is more related to ethological modelling than virtual characters, perhaps with the exception of team-based action games ("the weapons are behind the corner!"). However, someday, long-living agents inhabiting a large virtual societies in RPGs or yet-to-be-developed virtual worlds will need to share information for believability purposes; without sharing,

---

[2] We would like to thank to our colleague Jan Hajič for pointing us at this example.

[3] For present purposes, we conceive spatial memory systems as a part of episodic memory. Actually, spatial memory is a field of study of its own and its episodic nature is being discussed. For example, one major theory about the role of the hippocampus posits that its main function is spatial, while another theory argues for its role in processing of events. The neurobiological field seems to be interested in convergence of these two main threads of thinking (e.g. Eichenbaum 2004; Morris 2007, p. 581; Hassabis & Maguire, 2007).

[4] Humans are quite good in employing imagery to solve various problems ranging from path-planning to anticipating consequences of some situations to solving puzzles. Even though the nature of mental imagery is still hotly debated (see e.g. Pylyshyn, 2003; Kosslyn, 2006), it is clear that at least some of its aspects depend on the episodic memory system. Concerning the role of episodic memory in anticipation, see Zacks et al. (2007).

they would start to look as strange, uncommunicative individuals. Think of agents living in a closed area, such as a small city, that share information about a bizarre event that happened in past; an outsider should be recognised immediately for its unfamiliarity with the event. On a long time scale, in large long-running virtual worlds, we may even witness emergence of different "socio-cultural" groups of agents! Note that A7 skill departs from A2 in that A7 is oriented towards other agents while A2 towards users.

A8. Learning. Episodic memories can be exploited for the purposes of learning. For example, they can be used in an off-line manner during tuning of an agent's behaviour. Another possible use is for problem-solving; when an agent faces a problem, he can try to find whether he had not already solved a similar problem in the past and if he did, he can try to tackle the present problem in the way that worked then. Nuxoll (2007) points out similarity between these usages of episodic memories with case-based reasoning (Kolodner, 1993).[5]

Surely, we have not listed all possible skills capitalising on some facets of episodic memory, but the list is sufficient for the illustration that virtual characters may really need this memory. Arguably, the skills needed directly for interaction with users—A1, A2, A3—are most important for empathic characters. However, all other skills can be vital for some applications with empathic characters as well. Believability of an agent stems not only from user-agent interactions but also from the overall agent behaviour and agent-agent interactions while the agent is observed by the user.

Now, an FEM agent should possess all the skills at the same time. Can we imagine such an agent? What about an agent living in a magnificent yet-to-be-developed MMORPG or in a large social virtual world of the future (Goertzel, 2007)? The defining feature of this agent, besides her longevity, would be conversational abilities. This agent could have a regular "employment" in her virtual world, she could be a museum guide in a virtual museum for instance.

Well, but except of sci-fi examples, do we have really something? Unlikely. Most virtual characters would need some of these skills, but *not all* of them. Nevertheless, let us imagine that we have an FEM agent, that is, an agent with the A1–A8 skills; which requirements on the FEM architecture stem from having these skills?

2.2 Requirements on episodic memory

Let us start with real humans (see? the neuro-/psychology is coming...). Humans tend to segment the external flow into pieces organised around objects, actors, actions and the orders in which these elements combine to achieve specific goals: events (Nelson, 1986). Events take place in scenes: specific combinations of objects and/or situations at specific locations (Tversky et al., 2008). Events have a beginning and an end (Zacks et al., 2007), even though these may be sometimes fuzzy. Events can be either witnessed or communicated via language. It is these concepts—events, objects, actors, spaces, scenes, time, and language—we propose to design the FEM architecture around (Fig. 1):

B1. The notion of complex events. An FEM should support representation of complex real-world events that involve actors with human-level cognitive abilities. Such events have typically a nested structure – they can be logically decomposed to smaller events, until an atomic level is reached (Zacks & Tversky, 2001). This is in opposition to both merely physical events such as collision of two objects and laboratory events such as a presentation of a world list to memorise. This notion is demanded most by skills A1, A2, A6, A7, and A8.

B2. The notion of the time and the order. An FEM agent should be able to answer questions like "when something happened?", "what happened sooner?", or "what happened after something?". This feature is somehow demanded by all the skills; even for A4, the agent needs to compare the recency of memory records, and for A5, one can argue that some agents may need to remember past topologies ("there was a shop here, but now, there isn't").[6] But the notion of time has also other manifestations. For example, agents should use relative time concepts when speaking, such as "last summer," or "morning". This is not just an issue of mapping of absolute time units to a relative scale; relative notions are context depended—Monday morning is typically sooner than Sunday morning. Another sign of the time notion: agents should remember the course of interactions and be able to continue an interaction appropriately if interrupted (even today it may happen that when an RPG player returns to the pub he already visited, the virtual guests show no sign of remembrance him). These signs are most important for A1–A3. Yet another sign of the time notion: a long-living agent can be expected to adapt to a new way of life, e.g. after experiencing a change of a time-zone.

B3. The notion of objects and actors. An FEM agent should understand not only events in which she participates as the actor, but also events that she only observes. She has to understand who is the actor of an observed event (its causal factor) and what are the objects (the entities being manipulated with). Sometimes, there is no apparent causal factor, the case of "raining"; sometimes, there is a kind of "joint actor", the case of a dancing couple. This notion is most important for A1, A2, A3, A6, A7, and A8 skills. Sometimes, it may be sufficient to understand just affordances of an object ("this object can be used for this and that"), other times, features of an object and their changes may be needed as well ("the glass has been destroyed during that action"). Arguably, the latter is most vital for A8.

B4. The notion of space. An FEM should underpin many facets of spatial cognition. One of them is the topological knowledge about accessible locations, another is the awareness of the actual agent's surrounding, another is the long-term memory for positions of objects, yet another is the support for usage of linguistic terms describing spatial information, such as "left from" or "in front of me". Spatial skills go far beyond the A* and steering algorithms. Skills A4 and A5 benefit from the space notion; however, other

[5] Arguably, humans use past episodic memories to improve some of their problem-solving skills or semantic knowledge. However, the matters are not without controversy; for example, Tulving (2001) hypothesises that humans construct their semantic knowledge first and episodic memories second.

[6] Note that there is an inherent plausibility—folk psychology tension in the issue of timing. For example, people are known to be quite poor in dating, but may expect quite the opposite at the same time (Friedman, 1993).

skills may also require at least a rudimental understanding of space, including A1 and A3.

B5. The notion of scenes. Events do not take place in the abstract space, their stages are scenes; in a sense, scenes extends the notions of space, objects, and events.[7] While some scenes can be conceived as situation-based, it is the spatial facet that dominates in others. An example of the former is a "queue for something" while of the latter a "kitchen". Most skills require the notion of scenes, but while an FEM agent should have the ability to learn new scenes based on her interaction in the virtual world, most special-purpose agents act in limited domains, thus can be given the list of all the scenes *a priori*.

B6. The notion of language. In present context, the language is the medium for mediating knowledge about events (in a narrative-like way?). An FEM agent should be able not only to represent the flow of events based on what she directly perceives and feels, but also what someone tells her. This understanding is most important for the social skills A3 and A7. On the other hand, the FEM agent should be able to express her experience via language; the skills A1, A2, A7. Note that language can be actually used for building any declarative knowledge, including semantic knowledge.

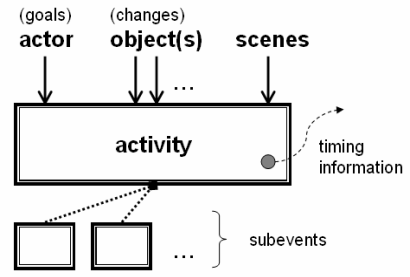

Fig. 1. The hypothetical unit around which episodic memories are organised (cf. Schank & Abelson, 1977; Zacks & Tversky, 2001).

2.3 The "definition" of the FEM

Now, the B-list from previous section is reasonably large and we can return to our original questions: 1) What are the features of the FEM? 2) What are FEM agents good for? We will first answer the former. Then, it will turn out, that we won't need to answer the latter anymore.

Notice that the list above tells us one important thing: there in no one to one mapping between the skills (As) and the requirements (Bs). For example, the skill A2 somehow underpins all the requirements; though some of them more (e.g. B1) while others less (e.g. B4). This and the fact that many agents need more than one skill (though not all of them) bring us to the hypothesis that many developers of today agents or agents to be built in the near future have to have similar, though not exactly same, requirements on episodic memory systems of their agents. Had every agent with a skill from the A-list demanded just one or two isolated mechanisms and, in addition, were these mechanisms different for every agent, it would make sense to develop these mechanisms during regular agent development, that is, to produce special-purpose solutions (which is what happens now for the few agents with episodic memory). However, it seems that this is not the case. Instead, there seems to be a large overlap of agent needs, hence there could be many (presently, non-existent) techniques that could be re-used. If this hypothesis is true, it would make sense to start a fundamental research program on generic episodic memory mechanisms, such mechanisms that can be picked by developers and customised for their agents similarly to how A* and steering techniques are now used. This research program would prevent developers to reinvent wheels as well as bring fruits of the integrative approach (when two mechanisms, such as a spatial memory and a memory for events, interact each other with, it is typically advantageous to start to investigate them together at some stage of progression; but this typically does not happen during regular development).

To sum up, it seems that there are strong reasons to start a research program, whose main goal would basically be:

> *to produce a bunch of ready-to-use mechanisms modelling some functional aspects of episodic memory for believable characters, capitalising on the integrative approach.*

The methodology of the program would be as follows: 1) to choose some mechanisms to investigate, 2) to investigate them in isolation *not* in the context of any specific application, 3) to wire them together again *not* in the context of any application, and to investigate how they communicate, influence each other, and hopefully produce emergent phenomena, 4) to customise this amalgamation or its parts for purposes of a specific application, 5) to add a new mechanism, returning somewhere between Stages (2) and (3). Of course, the selection made in Stages (1) and (5) should be well motivated, perhaps with the help of the A- and B-lists.

Now, we may return to Question (1). We have two possibilities how to define the FEM. First, we can say something like "the FEM is a bunch of memory systems that a) underpins the skills A1-A8 and b) is organised around the concepts B1-B6". Well, but we know that neither of the lists is definite. Imagine we define the FEM as suggested and an agent that needs the skills A1-A8 *plus* a new skill A9 will appear. This would be a silly situation: will we define something like FEM+? What to do if an A10 skill appear? It does not seem that this would be a useful definition.

But we now have also another possibility. Recall that the objective of the abovementioned research program is to produce a body of episodic memory mechanisms. We can define this body as the FEM. However, we think that this would be again a useless definition for this research will unlikely produce an outcome that will be fixed for eternity: the body of mechanisms would likely grow according to the needs of future agents.

What is the conclusion? We propose to resist the temptation to define the FEM for fruitlessness of this concept. Does this mean that the whole discussion was useless? It was not for two reasons. First, it helped us to isolate the A-list and the B-list, which are crucial for empathic characters. Second, it allowed us to formulate arguments for the advantage of integrative approach to the fundamental research on episodic memory for virtual agents. Given this conclusion, we should also resist the temptation to answer Question (2) for we have no definition of an FEM agent. However, this does not mean that the proposed research cannot produce many

[7] There are arguments based on fMRI experiments that the mental scene reconstruction is the key component process of various episodic and spatial memory abilities (Hassabis & Maguire, 2007).

interesting agents, as side-products in fact. If such an agent is developed and she finds no direct application, would it mean that the agent is useless? It won't for she would help to investigate the mechanisms of episodic memory, which will likely be directly applicable for another agents if the choices made during Stage (1) would be wise.

2.4 Some fruits of the integrative approach

We now illustrate two features of human episodic memory that goes across all or most of the points of the A-list and the B-list. Hence, it does not seem odd to investigate how these features can contribute to various mechanisms, not just to one special-purpose mechanism developed in isolation.

C1. Sparseness of encoding. Human episodic memory does not encode all available information. Some may not pass through the attention, some is likely encoded in an abstract way, without details. This applies for objects, spaces as well as events. For example, one may encode that an event happened at a scene "a place where I usually have breakfast" without encoding the colour of the table cloth. Or one may encode that he was "cooking", omitting the moment-by-moment course of the event (think of schemas (Bartlett, 1932)). Why episodic memory works in this way? The reasons seem to be "technical"; for instance, it is often argued that the following causes play their parts: the limited resources of our brains and the coding of the information in such a way that the information can be retrieved later easily after being cued.

C2. Forgetting and error susceptibility. Humans are not able to retrieve everything what they have encoded. Something can be retrieved only in the right context, something may be lost. Forgetting includes degradation of the content of episodes, spatial representation as well as temporal information. Its important feature is that it is *gradual* as opposed to *binary*. Different memories are forgotten in different speeds (likely based on their importance and emotional relevance). Similar memories can be eventually blended together. False memories can occur. Again, there are arguments that these "faults" are not faults but functional features of human memory (e.g. Schacter, 2002).

These points bring us to the widely accepted notion that human episodic memory has *reconstructive* nature, according to which the episodic memory is an active process of "constructing the past" that *engrave* memories and *reconstructs* them as opposed to merely *storing* them and *searching for* them in a database-like manner (e.g. Bartlett, 1932; Koriat & Goldsmith, 1996). Notice that the reconstructive nature underpins both C1 and C2.

Even though most present-day episodic memory agents have C1 feature, their memory systems tend to *store* everything what passed through a threshold mechanism of attention, and they typically do not employ forgetting or they use it in a simplified all-none fashion (see Strassner and Langer, 2005 for an exception). Although this approach is sufficient for most present-day applications (Ho & Watson, 2006), it may have two drawbacks from the long-term perspective. First, as suggested, the reconstructive nature of episodic memory is likely functional, it is technically advantageous. We believe that it will be easier to tackle some issues such as blending of episodes or limited computational resources when adopting the reconstructive perspective instead of the storage-based one. The second drawback is that storage-based memories are not psychologically plausible. However, it is not clear presently to which extent this is really an issue for believable agents need to be *folk* psychologically plausible, but not psychologically plausible. How exactly do humans expect episodic memory to behave? Here, we come to the second objective of the research program proposed above:

> *to investigate which features of agent episodic memory contribute to agents believability and which do not.*

Results of this line of research can also contribute to psychology. However, we have to develop the models first.

3. OUR AGENT

The purpose of this section is to review our on-going work on episodic memory for virtual characters which follows the research program defined in Sec. 2. For brevity, we will only sketch the main features of the model here. The model is detailed in the extended version of the paper (Brom & Lukavský, 2009), which also demonstrates benefits of the integrative research method taking various parts of the model as examples, and which gives some hints to empathic agents developers which parts of the model can be utilised in their applications.

Conceptually, the model integrates following parts: a visual short term memory, a long-term memory for "what-where" information, a life-long episodic memory, a component for timing, and a simple prospective memory. The action selection mechanism of the agent is a derivation of the BDI (Bratman, 1987). The agent features a simple valence-based emotion model. Presently, we have four independent implementations of various parts of the model, three of them employing a 2D grid world, the last one using a 3D world of the action game Unreal Tournament (Epic, 2004).

The key component of the model is the *long-term episodic memory* (LTEM), which has been already published in the paper that had proposed the notion of FEM (Brom et al., 2007). The LTEM represents what happened to the agent in the past, the flow of events. The memory is a hierarchical structure organised around tasks the agent can have in order to achieve some goals. The node of this structure resembles the unit on Fig. 1. The whole structure has some support in psychology (Zacks & Tversky, 2001). The fact that the tasks the LTEM stores have variable grain size allows for *gradual forgetting*: unimportant details of episodes can be forgotten. This memory has two mechanisms for storing timing information. One is based on time tags: when an event happens, an exact time information is added. This mechanism is simple to implement, but not plausible (Friedman, 1993). The second mechanism is a connectionist network that is able to a) acquire time concepts such as "morning" or "after lunch" based on the history of the agent's interaction, b) to represent timing information approximately, c) to gradually forget the timing information, d) to blend similar episodes that happened at different times.

One of the limitations of the LTEM is that it is not able to answer believably questions on positions of objects that are passive but whose locations can be changed by external forces. For this reason, the LTEM is intertwined with a memory for "what-where" information. This memory stores positional information in three frames of reference: egocentric, allocentric, and associative-based (the last one simply makes weighted associations between objects and places, estimating possible objects' locations). Our work in progress concerning this component is a mechanism that is able to learn notions of places based on where the agent lives such as "a

kitchen", "a corner in the kitchen", "a place in front of the monitor at the table" etc.

Together, the LTEM and the "what-where" memory underpin the skills A2 and A4. For example, if the agent is asked where are her glasses, she is able to answer: "likely at the bedside table, less likely next to the TV, and if they are not there, they might be somewhere in the living room or in the kitchen". If the agent is asked when she was gardening yesterday, she will answer "after lunch", not "from 2.13 to 4.12 p.m.".

There are several important points about this memory model. Most notably, the model is not a monolithic mechanism capitalising on a single representation, instead, it is a bunch of interconnected systems. Another thing is that even though it is not clear whether the agent featuring the whole memory can be directly utilised in a real-world application, the components of the memory can be. For example, virtual companions acting in the context of humans' flats (and in fact, robotic companions as well) can utilise the "what-where" map. Many long-living characters, such as RPG agents or storytelling agents, can use our LTEM, possibly with the timing mechanism. Finally, some of the mechanisms can be useful in other disciplines, for example in the subfield of psychology studying spatial cognition.

4. CONCLUSION OF EPISODE II

Episodic memory is one of the key components contributing to establishing the empathic relations with virtual agents, because it allows the user to understand better an agent's history, personality, and internal state. We have started with an idea that the *full* episodic memory might be an important, but yet-to-be defined, component of empathic agents. Now, our view is that it does not make a sense to define this component; instead, it is more fruitful to define a new research paradigm that investigates various episodic memory mechanisms capitalising on the integrative research method. The main goal of this paradigm is twofold: a) to develop a set of special purpose episodic memory techniques for agent developers, b) to investigate the plausibility—believability tension by evaluating the models with respect to real users. The paper also briefly reviewed our on-going work that can be regarded as pursuing this kind of research. Another contribution of this text is that it verbalised several fundamental skills of virtual characters that demand episodic memory and several notions around which episodic memory models should be organised.

To complete the picture it must be said that many issues concerning episodic memory have not been discussed here. For instance, how is the content of episodic memory related to the concept of self? (See Ho & Watson (2006) for more on this issue.) Would it be possible to generate the content of episodic memory automatically, e.g. using HTN-planning? There are many works on spatial cognition abilities in robotics; can we utilise some? Could a hardware chip for episodic memory be developed?

Exciting times seem to be at the horizon. Looking forward to Episode III.

5. ACKNOWLEDGMENTS

This work was partially supported by the Program "Information Society" under project 1ET100300517, and by the Ministry of Education of the Czech Republic (Res. Project MSM0021620838). The authors thank to students developing parts of the mentioned episodic memory model as their theses: Tomáš Korenko, Tomáš Soukup, Jan Vyhnánek, Jakub Kotrla, Klára Pešková, Rudolf Kadlec, and Ondřej Burkert.

6. REFERENCES

Baddley, A. 1986/1997. *Human Memory*. Psychology Press in Taylor & Francis Group.

Baddley, A., Conway, M., Aggelton, J. (eds) 2001. *Episodic Memory: New directions in research.* Oxford Uni. Press.

Baron-Cohen, S. 2003. *The Essential Difference.* New York: Basic Books.

Bartlett, F. 1932. *Remembering*. Cambridge Uni. Press.

Bratman M. E. 1987. *Intention, plans, and practical reason.* Cambridge, Mass: Harvard University Press.

Brainerd, C. J., Reyna, V. F. 2005. *The Science of False Memory*. Oxford Uni. Press.

Brom, C., Pešková, K., Lukavský, J. 2007. What does your actor remember. Towards characters with a full episodic memory. In *Proc. of 4th ICVS*, LNCS 4871. 89-101. Berlin, Springer-Verlag.

Brom, C., Lukavský, J. 2008. Episodic Memory for Human-like Agents and Human-like Agents for Episodic Memory. In *Tech. Reps FS-08-04 AAAI Fall Symposium BICA*, AAAI Press. 42 – 47

Brom, C., Lukavský, J. 2009. Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back – unabridged version. Tech. Rep. 3/2009. Dept Software Comp Sci Education. Charles University in Prague.

Burt, C.D.B., Kemp, S., Conway, M.A. 2003. Themes, events and episodes in autobiographical memory. In *Memory & Cognition* 31(2): 317-325

Cace, I., Bryson, J.J. 2007. Agent Based Modelling of Communication Costs: Why Information can be Free. In *Emergence and Evolution of Linguistic Communication*. 305-322

Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P. W., Ho, S. 2008. Long-term affect sensitive and socially interactive companions. In 4th *Int. Workshop on Human-Computer Conversation*.

Companions: Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet. 2006. The European Community's Seventh Framework Programme. URL: http://www.companions-project.org/ [12.2.2009]

Conway, M. A. 2005. Memory and the self. In *Jn. of Mem. Lang.* 53: 594-628

Dias, J., Ho, W.C., Vogt, T., Beeckman N., Paiva, A., Andre, E. 2007. I Know What I Did Last Summer: Autobiographic Memory in Synthetic Characters. In *Proc. of ACII*. 606–617. Berlin, Springer-Verlag.

Dodd, W. 2005. The design of procedural, semantic, and episodic memory systems for a cognitive robot. Master thesis. Vanderbilt University, Nashville, Tennessee.

Eichenbaum, H., Cohen, N. J. 2001. *From conditioning to conscious recollection.* Oxford Uni. Press

Eichenbaum, H. 2004. Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron* 44: 109-120.

Epic Epic Games: UnrealTournament 2004. http://www. unrealtournament.com [12. 2. 2009]

Friedman, W.J. 1993. Memory for the Time of Past Events. In *Psychol Bull* 113(1): 44 – 66

Goertzel, B. 2007. AI Meets the Metaverse: Teachable AI Agents Living in Virtual Worlds. URL: http://lifeboat.com/ex/ ai.meets.the.metaverse [12.2.2009]

Hassabis, D., Maguire, E. A. 2007. Deconstructing episodic memory with construction. In *Trends Cogn Sci* 11(7): 299-306

Ho, W. C., Watson, S. 2006. Autobiographic knowledge for believable virtual characters. In *Proc. of Intelligent Virtual Agents*, LNCS 4133. 383-394. Berlin, Springer-Verlag.

Ho, W., Dautenhahn, K., Nehaniv, C. 2008. Computational Memory Architectures for Autobiographic Agents Interacting in a Complex Virtual Environment. *Connection Science*. In press.

Isla, D., Blumberg, B. 2002. Object Persistence for Synthetic Creatures. In *Proc. AAMAS'02*, 1356-1363

Johnson, W. L. 1994. Agents that learn to explain themselves. In *Proc. of the 12th Nat. Conf. on Artificial Intelligence*, 1257-1263. AAAI Press.

Kandel, E. 2001. The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses. In *Science* 294. 1030-1038

Kolodner, J. 1993. *Case-Based Reasoning*. Morgan Kaufmann.

Kopp, S., Gesellensetter, L., Krämer, N. C., Wachsmuth, I. 2005. A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In *Proc of IVA*, LNCS 3661. 329-343. Berlin, Springer-Verlag.

Kosslyn, S. M. 2006. *The Case for Mental Imagery.* Oxford Uni. Press.

Koriat, A. & Goldsmith, M. 1996. Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. In *Behavioral and Brain Sciences* 19(2): 167-228

Krichmar, J. L., Seth, A. K., Douglas, A., N., Fleischer, J. G., Edelman, G. M. 2005. Spatial Navigation and Causal Analysis in a Brain-Based Device Modelling Cortical-Hippocampal Interactions. In *Neuroinformatics* 3(3): 197-221

Kuipers, B. 2000. The Spatial Semantic Hierarchy. In *Artificial Intelligence* 119: 191-233

Lim, M. Y. 2007. *Emotions, Behaviour and Belief Regulation in An Intelligent Guide with Attitude.* Ph.D. diss. School of Mathematical and Computer Sciences, Heriot-Watt University.

Loftus, E. F. 1979. *Eyewitness testimony*. Harvard Uni. Press.

Loyall, B. A. 1997. *Believable Agents: Building Interactive Personalities*. Ph.D. diss. Carnegie Mellon University.

McNaughton, B.L. et al. 2003. Off-line reprocessing of recent memory and its role in memory consolidation: a progress report. In *Sleep and Brain Plasticity.* 225-246. Oxford Uni. Press.

Miyake, A., Shah, P. eds. 1999. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press.

Morris, R. 2007. Theories of hippocampal function. In *The hippocampus Book*, 581-694. Oxford University Press.

Nelson, K. (ed.) 1986. *Event knowledge: Structure and function in development.* Lawrence Erlbaum.

Noser, H., Renault, O., Thalmann, D., Magnenat-Thalmann, N. 1995. Navigation for Digital Actors based on Synthetic Vision, Memory and Learning. *Computer and Graphics*, 19(1): 7-19

Norman, K. A., Detre, G. J., Polyn, S. M. 2008. Computational models of episodic memory. In *The Cambridge Handbook of Computational Cognitive Modeling*. In press.

Nuxoll, A. 2007. *Enhancing Intelligent Agents with Episodic Memory*. Ph.D. diss. The University of Michigan.

Paiva, A., Dias, J., Sobral, D., Woods, S., Hall, L. 2004. Building Empathic Lifelike Characters: the proximity factor. In *Proc of Workshop on Empathic Agents*, AAMAS'04.

Peters, C. 2006. Designing Synthetic Memory Systems for Supporting Autonomous Embodied Agent Behaviour. In *Proc. RO-MAN06*. 14-19.

Pylyshyn, Z. 2003. Return of the mental image: are there really pictures in the brain? In *Trends in Cogn Sci* 7(3): 113-118

Rickel, J., Johnson, W. L. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *App. Artificial Intelligence* 13(4-5): 343-382

Schacter, D. L. 2002. *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Mariner Books.

Shipley, T. F., Zacks, J. M. (eds) 2008. *Understanding Events*. Oxford Uni. Press.

Schank, R. C., Abelson, R. P. 1977. *Scripts, plans, goals, and understanding.* Hillsdale, N.J: L. Erlbaum Associates.

Squire, L. R., Zola-Morgan, S. 1991. The medial temporal lobe memory system. In *Science* 253: 1380-1386

Strassner, J., Langer, M. 2005. Virtual humans with personalized perception and dynamic levels of knowledge. *Comp. Anim. Virtual Worlds* 16: 331-342

Thomas, R., Donikian, S. 2006. A spatial cognitive map and a human-like memory model dedicated to pedestrian navigation in virtual urban environments. In *Proc. Spatial Cognition V*, LNCS 4387. 421-436. Berlin, Springer-Verlag.

Tulving, E. 2001. Episodic memory and common sense: how far apart? In *Episodic Memory: New directions in research.* Oxford Uni. Press. 269-288.

Tulving, E., Donaldson, W. 1972. *Organization of memory*. New York: Academic Press.

Tversky, B., Zacks, J. M., Hard, B. M. 2008. The Structure of Experience. In *Understanding Events*. Oxford Uni. Press.

Wagenaar, W. A. 1986. A study of autobiographical memory over six years. In *Cogn Psychol* 18: 225-252

Williams, H. L., Conway, M. A., Baddley, A. D. 2008. The Boundaries of Episodic Memories. In *Understanding Events*. Oxford Uni. Press.

Zacks, J. M., Tversky, B. 2001. Event Structure in Perception and Conception. In *Psychol Bull* 127(1): 3-21

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., Reynolds, J. R. 2007. Event Perception: A Mind-Brain Perspective. In *Psychol Bull* 133(2): 273-293