

# I've Been Here Before! Location and Appraisal in Memory Retrieval

To be published in Int. Conf. on Autonomous Agents and Multiagent Systems  
(AAMAS 2011)

Paulo F. Gomes  
INESC-ID and Instituto  
Superior Técnico  
Av. Prof. Dr. Aníbal Cavaco  
Silva  
2744-016 Porto Salvo,  
Portugal

Carlos Martinho  
INESC-ID and Instituto  
Superior Técnico  
Av. Prof. Dr. Aníbal Cavaco  
Silva  
2744-016 Porto Salvo,  
Portugal

Ana Paiva  
INESC-ID and Instituto  
Superior Técnico  
Av. Prof. Dr. Aníbal Cavaco  
Silva  
2744-016 Porto Salvo,  
Portugal

## ABSTRACT

The objective of our current work was to create a model for agent memory retrieval of emotionally relevant episodes. We analyzed agent architectures that support memory retrieval realizing that none fulfilled all of our requirements. We designed an episodic memory retrieval model consisting of two main steps: *location ecphory*, in which the agent's current location is matched against stored memories associated locations; and *recollective experience*, in which memories that had a positive match are re-appraised. We implemented our model and used it to drive the behavior of characters in a game application. We recorded the application running and used the videos to create a non-interactive evaluation. The evaluation's results are consistent with our hypothesis that agents with memory retrieval of emotionally relevant episodes would be perceived as more believable than similar agents without it.

## Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—*Intelligent agents*

## General Terms

Theory, Experimentation

## Keywords

Modeling cognition and socio-cultural behavior, Affect and personality, Virtual character modeling and animation in games, education, training, and virtual environments

## 1. INTRODUCTION

Towards the end of the 20th century, computer scientists in the field of autonomous agents, began to analyze how the

**Cite as:** I've Been Here Before! Location and Appraisal in Memory Retrieval, Paulo F. Gomes, Carlos Martinho and Ana Paiva, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.

Copyright © 2011, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

artistic principles of animated characters could be used to design believable agents. For instance, Bates' work in the OZ Group [2] was inspired by Thomas and Johnston's *The Illusion of Life: Disney Animation* [17]. Two of the key ideas guiding Bates were: an agent's emotional state must be clearly defined; and the agent's actions must express what it is thinking about and its emotional state. Loyall [9], also working in the OZ Group, further dissected the definition of agent believability, proposing among others, the following requirements: ability to grow and the behavior changing according to different situations. Consequently the agent's behavior should in principle reflect what it has lived. This idea is also consistent with Ortony's believability definition [12]. He considers that the evaluation perspective and behavior displayed by an agent should be coherent across different types of situations and over the agent's experience. Finally, although in many cases not being strictly believable agents, believable characters in video games share characteristics with the former. Rollings and Adams [15] proposed that believable video game characters should grow with the game story and overall game experience (pp. 134–135).

The ideas presented point to one architectural element: memory. In particular, personal memories concerning emotionally relevant episodes to the agent, as emotion is also a crucial element for agent believability. In humans, this type of memories may be considered episodic [18]: memories that refer to personal experiences that are linked with a specific time and place (e.g. I left my keys on top of the fridge in the kitchen yesterday night). Episodic memories and semantic knowledge (general knowledge about the world and facts one knows) are often analyzed together as autobiographic memory [1]: episodic memories can be combined together, or even generalized to semantic knowledge.

Focusing on episodic memories, Tulving [18] stated that they enable humans to do mental time travel, that is, to re-live past experiences. This re-experience takes place during retrieval. Retrieval of episodic memory involves the interaction between a memory trace, "a physical representation of a memory in the brain" [4], and a retrieval cue, a stimulus that can be either internal or external [16] (e.g. smells). In [19] episodic memory retrieval is described as a two staged process: ecphory and conversion. "Ecphory is a process by

which retrieval information provided by a cue is correlated with the information stored in an episodic memory trace”. The product of ephory is a set of pairs of highly correlated cues and traces. These pairs are then converted into a recollective experience (conversion). It is through the recollective experience that a person is able to relive a past event [18], although typically not as intensively as before [11].

Motivated by the definitions of believability and by the human memory theory principles presented, we created a model for episodic memory retrieval in believable agents. We model retrieval of emotionally relevant episodes through ephory and recollective experience (with re-appraisal of past events). We believe that this model can promote the perceived believability of agents. Furthermore, the model can be used to enhance non-player characters’ behavior in video-games, by making it dependent on their personal experiences.

## 2. RELATED WORK

Agent architectures that support autobiographic memories have been proposed in several previous works, however modelling episodic memory retrieval for believable agents is not an extensively debated topic.

Ho, Dautenhahn and Nehaniv have developed several autobiographic memory architectures in the context of artificial life [6]. In them, memories are paths to resources, and storage is triggered by a timer or by an event. Retrieval happens when a resource is needed and the retrieval process consists of reconstructing previously walked paths. Although the model supports retrieval of personal information (the agent’s paths), this retrieval is not an emotional experience. Moreover, the work focuses much more on the agent’s survival skills, than on believability characteristics.

In FATiMA [5] there is a greater concern with believability, with emotion and personality being central concepts. The system supports both appraisal and autobiographic memories. Furthermore, these memories contain the agent’s emotional reaction to the events. Nevertheless, memory retrieval is not presented as an emotional appraisal process: memories are simply retrieved when the agent wishes to summarize his life story.

In [7] memories are combined and encode both goals and event coping strategies. Autobiographic memories are used for extracting goals, verifying if they have been achieved and choosing a reaction strategy to an event. The model is, however, clearly more directed to the semantic part of autobiographic memory, than to episodic memory.

Brom et al [3] proposed an agent architecture that supports episodic memory retrieval. In long-term memory, memories are structured as trees of performed tasks. These tasks are removed from long-term memory according to a forgetting mechanism that takes into account, among other things, the time passed since the task was performed. Despite all its features, memory retrieval is described as a data base process, and not as an emotional experience.

In brief, modelling ephory and modelling an emotional recollective experience are relatively unexplored subjects in the analyzed work. We will delve into them in the next section.

## 3. MODEL

We have developed a model for agent episodic memory

retrieval (see Figure 1) with two main steps: *location ephory* and *recollective experience*.

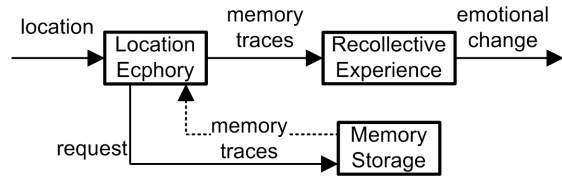


Figure 1: Episodic memory retrieval

### 3.1 Location Ephory

Our model is motivated by the idea that humans retrieval process results from the interaction between memory traces and retrieval cues (stimuli) [16]. If a person is exposed to stimuli similar to the ones he, or she, was exposed during the occurrence of an event that is stored in memory, these stimuli can act as retrieval cues for that memory.

Consider the following situation: an individual A passing by the spot where she was first kissed. As individual A passes by, she might smell the sent of near flowers, be again exposed to the colors of the garden, gaze at the mountain landscape, feel the crunchy texture of the ground. All of these external stimuli act as retrieval cues, and individual A remembers her first kiss. Note that the mentioned stimuli are perceived in the garden. Hence, instead of saying that the individual stimuli elicit the kissing memory, one can say that the garden’s stimuli elicited the episodic memory. In the end, *the garden’s location is acting as an indirect retrieval cue for the memory*.

Of course if the garden had been replaced by a parking lot, and the view was now hidden by a shopping mall, the retrieval cues would be absent, and consequently the location could hardly be seen as an indirect retrieval cue for the episodic memory. Thus, the exposed situation as a whole shows that *locations can be interpreted as indirect memory retrieval cues when they have not changed dramatically*.

We can translate our intuition, by defining that location ephory selects memory traces whose connected event occurred close by the location where the agent currently is. It is a simplification of the generic ephory: on one hand it replaces direct stimuli input by physical locations; on the other hand, it only accounts for retrieval of a memory trace when passing by the location where the memory trace’s past event took place.

In spite of its limitations, from an engineering perspective, location ephory is much less demanding on the sensor detail of a synthetic autonomous agent. Agents just need to be able to approximate their current physical location. They do not need to have a wide range of simulated sensors covering smell, sights, sounds, colours, etc. The ability to approximate a current physical location is much more common in agents than detailed simulated perception. Therefore we believe that location ephory can be integrated into a wider range of agent architectures than a more generic ephory model.

### 3.2 Recollective Experience

After the traces are selected by location ephory, there still needs to be a recollective experience. According to Tulving [18] episodic memories allow humans to relive past

experiences. Analogously, if we consider that an agent appraises an event when it first experiences it, then when it “relives” the event we propose a second appraisal should take place. Therefore, when a memory trace is selected by location ephory the event that is linked to that memory trace is appraised. Hence, the recollective experience will essentially be an appraisal process.

Before we further describe the recollective experience, we need to define the concept of emotional reaction, emotion and emotional state. These definitions are inspired in the OCC model [13] and on FAtiMA [5]. We start by laying down a background scenario that will serve to exemplify the emotion definitions.

Two agents (meemo 1 and meemo 2) are moving in a tunnel. Meemo 1 and meemo 2 are friends. Meemo 1 witnesses meemo 2 falling in a deadly trap. Meemo 1 evaluates this event as undesirable for meemo 2 and also as undesirable for itself (as meemo 2 was its friend). Meemo 1 will have an emotional reaction to the event.

In our model an *emotional reaction* is a quantified evaluation of an event, defined by a pair  $\langle AV, E \rangle$  in which:

- $AV$  contains the set of appraisal values, two of which are desirability-for-self and desirability-for-other. Each appraisal variable represents an evaluation of the event through a specific perspective of the agent. Desirability-for-self represents the extent to which an event enables, or hinders, the achievement of a personal goal. Desirability-for-other is the inferred desirability of an event for another individual. In our example, meemo 1 might have as a goal “stay alive” which will lead to a low value of desirability-for-self. Additionally meemo 1 can have a goal “meemo 2 stay alive” which leads to an even lower value of desirability-for-other.
- $E$  specifies the event that generated the reaction. In the example, this element might have information such as “meemo 2 fell in trap located in tunnel on spot b3”. We use the term event as a generalization of the OCC’s appraisal evaluation focus: on consequences of an event, on the agency element of an event, or on an object of an event.

We define *emotion* as a valanced evaluation of an event described as a 4-tuple  $\langle E, ET, EI, V \rangle$  in which:

- $E$  contains information about the event that elicited the emotion (e.g. “meemo 2 fell in trap located in tunnel on spot b3”).
- $ET$  specifies the emotion type according to the OCC model [13] (e.g. pity).
- $EI$  specifies the current intensity scalar value (non-negative).
- $V$  specifies the valence of the emotion (positive or negative). The valence is directly dependent on the emotion type. For example, joy emotions are positively valanced and pity emotions are negatively valanced.

An *emotional state* is defined by a 2-tuple  $\langle AE, M \rangle$  in which:

- $AE$  contains the set of emotions the agent is currently feeling.

- $M$  specifies the mood value. Mood is a bounded scalar value that represents the agent’s overall emotional state valence. Low values represent a bad mood and high values represent a good mood. For example, meemo 1 learns how to detect traps, causing it to feel joy, and in turn rising its mood. Shortly afterwards it detects a trap and feels pride, causing its mood to rise even higher.

We can now proceed with the model’s description. The recollective experience process flow has three main steps:

1. Generating emotional reactions from events.
2. Generating emotions from emotional reactions.
3. Integrating generated emotions into the emotional state.

Extensive work has been done regarding all these steps, being FAtiMA [5] and Ema [10] examples of this. For the recollective experience one just needs to use a model such as the ones just mentioned. The past event information is extracted from the selected memory trace and then this information is fed into a generic appraisal module<sup>1</sup>. Our model ties in with the OCC model [13], as it specifically refers that appraised events can be in the recent or remote past (pg. 86).

However, if we consider a generic appraisal module, some modifications need to be made. Following the view that a person can relive a past event as an observer or as an actor [11], agents will be able to do the same. Different architectures of appraisal use different structures for creating emotional reactions (construal frames, plans, reactive rules, etc), and these structures can change over time. When re-appraising an event the agent will be able to evaluate it according to its current evaluation structures (as an observer of its “past-self”), or use the emotional reaction to the event when it first occurred (as an actor in the event).

After emotional reactions to events have been created (*step 1*), they can be used to generate emotions (*step 2*). In a generic appraisal module, the only change that needs to be made, is to decrease the intensity of emotions, or of potential emotions, when they are generated by re-appraisal of past events. With this decrease we try to encode the idea that memory retrieval is, in general, a less intense experience than the original one [11]. *Step 3* of a generic appraisal system does not need to be modified when the system is used to create a recollective experience.

### 3.3 Memory Storage

In general, each emotion that was successfully generated is passed to memory storage, together with the event that caused the emotion. Choosing to store emotion eliciting events is supported by research stating that in humans emotions drive event focus and consolidation [14], and that emotion arousal extends the durability of memories [11]. However, if the emotion was generated due to a retrieval event, no memory trace is stored. This choice was made to avoid recursive memory retrieval.

Memory storage creates an episodic memory trace as a 5-tuple  $\langle Pp, D, T, Er, Em \rangle$  in which:

<sup>1</sup>We will use the term *retrieval event* to refer to a past event that will be re-appraised.

- $D$  contains a description of the event including where it occurred (e.g. companion fell in trap at location (30,60)).
- $T$  defines the time stamp when the event started.
- $Er$  specifies the emotion reaction to the event.
- $Em$  specifies the emotion elicited by the appraisal of the event.

Memory traces are initially stored in a short-term memory storage (STM), and after a few seconds are passed to the long-term memory storage (LTM). It should be noticed, that only events that elicit emotions are stored at all, hence we filter memory traces before they go to STM.

Additionally, when a memory trace is selected by ephory, it passes from LTM to STM. Retrieval abstractly represents passing memories from long-term memory to short-term memory. Consequently, if they are already in short-term memory, they should not be retrieved. Hence ephory only selects memory traces that are in LTM, and ignores memory traces in STM.

As a final remark it should be noted that no model for memory forgetting will be presented. Our research focus is on episodic memory retrieval, consequently only the memory storage elements strictly relevant to the retrieval process are described. Nonetheless, a forgetting mechanism similar to the one presented in [3] could be easily adapted for this purpose.

## 4. IMPLEMENTATION

Having defined a model for episodic memory retrieval in the previous section we will now describe how it was implemented. First of all we present an overview of the agent architecture (schematically represented in Figure 2). The *Location Ephory* module is responsible for constantly trying to match the agent’s current location with stored memory traces. If there is a match, the memory trace’s event is fed into the *Appraisal* as a retrieval event. The *Appraisal* acts as a *Recollective Experience* enabling retrieval events to be re-experienced. In parallel, non-retrieval events (present events), when generated, are also fed into *Appraisal*. All events are appraised and, as a consequence the emotional state may be changed. If the emotional state is changed due to a non-retrieval event, the causing event is stored as a memory trace in *Memory Storage*. Meanwhile, the *Behavior* uses the emotional state, and memory traces from *Memory Storage*, to determine which actuators should be activated.

### 4.1 Events

In the architecture’s overview, we mentioned that all events are fed to the *Appraisal*. These events are generated either by sensors (non-retrieval events) or by the *Location Ephory* (retrieval events). Non-retrieval events have two main parameters indicated in *Parametrization 1*.

*Parametrization 1. Non-Retrieval Event*

**type** Enumerate representing the type of the event. In the application it is assumed that there is a finite number of event types.

**location** If the event took place at a specific point in space, it will have the world coordinates of that point (e.g. if

an agent finds a raspberry bush, the location of this event could be the exact coordinates of the bush). If however, the event’s action is spread trough an area, the location will be the world coordinates of a point representing the event’s action center (e.g. if an agent performs a dance in an area, the location of this event can be the centroid of that area).

There is a special type of non-retrieval events called *witness events*. In witness events the agent is not an agency element of the event, that is, the agent’s actions are not directly causing the event. Witness events have type *EventWitness* and have an additional parameter (*witnessed event*). This parameter represents the event being witnessed by the agent. All events, including witness events, can elicit emotions in the agent. Events and caused emotions are stored together as memory traces.

### 4.2 Memory Encoding and Storage

A memory trace has only three parameters as presented in *Parametrization 2*. There is no emotion reaction parameter because we only implemented the recollective experience as an observer, hence the emotion reaction was not necessary.

*Parametrization 2. Non-Retrieval Event*

**event** which the memory is about.

**emotion** caused by the event.

**time stamp** when the event started or when it was retrieved for the last time (details presented bellow).

We conceptually separate memory traces in long-term memory (LTM) from memory traces in short-term memory (STM). A memory trace is considered to be in STM if the difference between its time stamp (TS) and the current time is smaller than *short term memory duration* (Equation 1). Short term memory duration (*stmd*) can be parameterized and has as default value 20 seconds. This choice is inspired by the idea that in humans information is kept in short-term memory for up to 20, to 30 seconds, if no rehearsal takes place [4](pg. 696).

$$CurrentTime() - TS(memory\ trace) < stmd \quad (1)$$

Memory traces that do not verify this condition, are considered to be in LTM. When created, a memory trace starts by being in STM. While in STM a memory trace can not be selected for ephory. After the short term memory duration has elapsed, it is considered to be in LTM. If the memory trace is selected by *Location Ephory*, its timestamp is updated to the current time, hence the trace passes again to STM. Another short term memory duration will have to pass before the memory trace is in LTM again, and can be selected once more by *Location Ephory*.

### 4.3 Location Ephory

At each time step, *location ephory* matches all memory traces in *Memory Storage* against the agent’s current location. If the euclidean distance (*ED*) between the agent’s current location, and the memory trace’s event location, is smaller than *location ephory distance* (*led*), parameterizable in a configuration file, there is an ephory match.

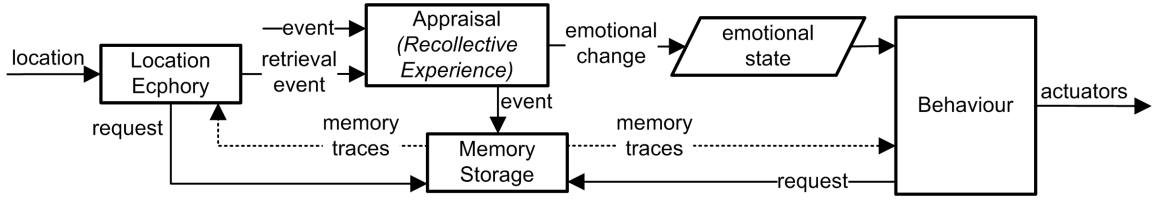


Figure 2: Agent Architecture

$$ED(L(agent), L(E(memory\ trace))) < led \quad (2)$$

Consequently, when an agent is in the close proximity of a location where an event took place, and that event is stored in the agent’s LTM (through a memory trace), memory retrieval of that event is triggered. In this process, more than one memory trace may be selected, because several memories can be linked with past events that occurred close to where the agent is. For each memory trace that was selected a retrieval event is created.

Besides the parameters previously presented for non-retrieval events, retrieval events have an additional one: retrieved event. The retrieved event is the event parameter of the memory trace that was selected. Furthermore, the type parameter is set to “Retrieval” and the location parameter is set to the location value of the respective retrieved event parameter. Generated retrieval events are fed into the Recollective Experience (Appraisal). Additionally, matching memory traces get their timestamp updated to the current time.

Consider the following example in which the location ecphory distance was set to 2 meters and locations are defined in a two dimensional space. An agent  $a1$  has three memory traces in Memory Storage:  $m1$ ,  $m2$  and  $m3$ . Their events are respectively  $e1$ ,  $e2$  and  $e3$ , and these events’ locations are  $l1$ ,  $l2$  and  $l3$ . The agent is currently at location  $la1$ . All locations are schematically represented in Figure 3. Additionally, we know that  $m2$  and  $m3$  are in LTM while  $m1$  is STM. In this situation there would be an ecphoric match for  $m2$  because  $l2$  is closer than 2 meters from  $la1$  and  $m2$  is in LTM. There would be no ecphoric match for  $m3$  ( $l3$  is further than 2 meters from  $la1$ ) nor for  $m1$  ( $m1$  is in STM). Only  $m2$  will be selected for Recollective Experience. Hence, a retrieval event  $re$  will be generated with retrieved event parameter set to  $e2$  and its location parameter set to  $l2$ . Retrieval event  $re$  will then be fed into Appraisal. Meanwhile, as  $m2$  was selected, it passes to STM, and consequently will not be able to be selected again for Recollective Experience for the duration of short term memory duration.

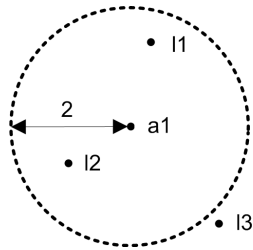


Figure 3: Location Ecphory Example

## 4.4 Appraisal

As previously mentioned, the Appraisal is used to evaluate present events as well as re-experience past ones. To develop it, we started by translating from the Java programming language to C++ the reactive appraisal part of FAtiMA’s implementation [5], adapting it when necessary. For instance, we changed it so it would treat differently retrieval events and non-retrieval events. To describe all the Appraisal’s elements, we will follow the steps defined in the model for a generic Recollective Experience process flow (see Section 3.2).

### 4.4.1 Recollective Experience - step 1

Appraisal receives retrieval events from Location Ecphory, and non-retrieval events generated by sensors. It starts by using these events to produce emotional reactions. An emotional reaction has the parameters presented in Parametrization 3.

*Parametrization 3.* Emotional Reaction Parameters

**desirability for self** Integer varying between -10 and 10 (except 0), or *null appraisal value* (integer not in this range). A negative value indicates that the event hinders the achievement of an agent’s goal, and a positive value indicates that the event enables the achievement of an agent’s goal. *Null appraisal value* indicates that the event has no effect on the agent’s goals.

**desirability for other** Same as desirability for self but in regard to other agents’ goals.

**praiseworthiness** Similar to desirability for self but concerning violation, or uphold, of agent’s standards.

**event** Event that caused the emotional reaction.

Emotional reactions are generated from events using reaction rules. A reaction rule has the same parameters as an emotional reaction, however its event does not have a defined location. Each agent has a set of reaction rules. Each received event is matched against all reaction rules of this set. Matching consists of comparing the event with the reaction rule’s event. In turn, comparison between two events is done using a function whose result values vary between 0 (no match) and 10 (total match). Two events of different types have a comparison value of 0. Two events with all parameter values equal have a comparison value of 10.

When the comparison value between an event and the reaction rule’s event is positive an emotional reaction is generated. This emotional reaction has the same parameter values for desirability for self, desirability for other and praiseworthiness as the reaction rule, and the event parameter is set to the event that was matched with the reaction rule.

Ultimately, reaction rules serve to implicitly represent the agent's goals.

There is one exception to the generic matching process described, that concerns reaction rules for retrieval events (that we will name *retrieval reaction rules*). The idea behind appraising retrieval events, as described in the model, is that by doing so the agent is able to relive past events, similarly to episodic memory retrieval in humans [18]. We have implemented this by creating an additional reaction rule for each reaction rule containing a non-retrieval event. The new reaction rule has the same desirability for self, desirability for other and praiseworthiness values as the original one. However its event parameter is a retrieval event, that in turn has its retrieved event parameter set to the event of the original reaction rule.

Regarding the matching process, if a reaction rule's event is a retrieval event (retrieval reaction rule), and the event to be matched is not of type "Retrieval", the comparison value is 0, as described in the generic case. However, if the event to be matched is of type "Retrieval" a second comparison must take place: the retrieval event parameter of the reaction rule's event must be compared with the retrieval event parameter of the event to be matched. The result obtained for the retrieval event parameters is used for matching the reaction rule to the original event. With this we model the agent appraising a past event according to its current appraisal structures (in this case the reaction rules).

#### 4.4.2 Recollective Experience - step 2

Returning to the architecture's description, an emotional reaction is generated when an event and a reaction rule match. Generated emotional reactions are used to create potential emotions. This process starts *step 2* of the Recollective Experience model described in Section 3.2. A potential emotion has the parameters presented in Parametrization 4.

##### Parametrization 4. Potential Emotion Parameters

**event** Event that generated the emotional reaction.

**base potential** Scalar between 0 and 10 that represents the potential intensity of the emotion.

**type** Enumerate that represents the emotion type according to the OCC model [13]. The implementation generates potential emotions of the following types: Joy, Distress, HappyFor, Resentment, Gloating, Pity, Pride, Shame, Admiration and Reproach.

**valence** POSITIVE if the emotion type is Joy, HappyFor, Gloating, Pride and Admiration. NEGATIVE for all other types.

An emotional reaction can generate a maximum of three potential emotions because each emotional reaction can elicit at most one emotion of each of the following categories of the OCC model [13]: focus on consequences of events for others (HappyFor, Resentment, Gloating and Pity), focus on consequences of events for self when prospects are irrelevant (Joy and Distress) and focus on actions of agents (Pride, Shame, Admiration and Reproach). We will name these three categories *focus on others*, *focus on self* and *focus on actions*, respectively.

If the *desirability for self* of an emotional reaction is not *null appraisal value*, a potential emotion of the *focus on self*

category will be generated. If *desirability for self* is negative the potential emotion's type will be Distress, if it is positive the emotion type will be Joy. In both cases the base potential (*bp*) will be the absolute value of *desirability for self* ( $bp = |desirability\ for\ self|$ ). In this category, as well as in the other two, the event parameter is always set to the emotional reaction's event.

If both *desirability for self* (*dfs*) and *desirability for other* (*dfo*) of the emotional reaction are different from *null appraisal value*, a potential emotion of the *focus on other* category will be generated. The base potential in this case is given by the expression  $\frac{|dfs|+|dfo|}{2}$ . The type of the potential emotion is defined according to the values of *desirability for self* and *desirability for other*:

- *HappyFor*:  $dfs > 0$  and  $dfo > 0$ ;
- *Gloating*:  $dfs > 0$  and  $dfo < 0$ ;
- *Resentment*:  $dfs < 0$  and  $dfo > 0$ ;
- *Pity*:  $dfs < 0$  and  $dfo < 0$ ;

Finally, if the *praiseworthiness* (*pw*) of the emotional reaction is different from *null appraisal value*, a potential emotion of the *focus on actions* category will be generated. The base potential will be the absolute value of *praiseworthiness* ( $bp = |pw|$ ). The type of the potential emotion is defined according to the values of *praiseworthiness* and to the emotional reaction's event type:

- *Pride*:  $pw > 0$  and  $event\ type \neq EventWitness$ ;
- *Admiration*:  $pw > 0$  and  $event\ type = EventWitness$ ;
- *Shame*:  $pw < 0$  and  $event\ type \neq EventWitness$ ;
- *Reproach*:  $pw < 0$  and  $event\ type = EventWitness$ ;

In witnessed events the agency element of the event is not the agent, therefore potential emotions caused by an emotional reaction to them should be directed outwards (Admiration or Reproach) and not inwards (Pride or Shame).

After a potential emotion is created, independent of which category it belongs to, its base potential is recalculated if the event parameter is a retrieval event. The new base potential is determined by Equation 3, in which *memory retrieval intensity bias* is a configurable positive value smaller than one and *oldBP* is the base potential before recalculation.

$$oldBP \times memory\ retrieval\ intensity\ bias \quad (3)$$

By using such an expression the base potential of potential emotions generated from emotional reactions to retrieval events, will be smaller in comparison to ones for which the event is a non-retrieval event. Consequently, when an agent reappraises a past event, the base potential of the corresponding potential emotion will be smaller than the base potential of the potential emotion originally generated when the past event was appraised. This formula tries to encode the idea, described in the model, that the memory retrieval's experience is, in general, less intense than the original experience [11].

For the remaining emotional process we only did minor changes to FATiMA's implementation [5]. Therefore we will only describe it in brief.

Two other factors, besides the previously mentioned, contribute for emotions’ intensities: mood and emotion thresholds. If an agent is in a good mood, positive emotions will be favored and negative ones lessened in intensity. A negative mood has the opposite effect. An emotion threshold, on the other hand, defines a minimum value an emotion has to have in order to be activated. This value is subtracted to the emotion’s base potential when calculating its final intensity. Thresholds are agent specific and emotion specific. They can be seen as the resistance an agent has to a certain emotion, and be used to model personality.

#### 4.4.3 Recollective Experience - step 3

After the final intensities are calculated, the emotions are integrated into the emotional state, that consists of the already mentioned mood value and of a set of active emotions. All emotions are added to the set, with positive emotions increasing the mood value and negative ones decreasing it. Note that emotion intensities, as well as the mood’s absolute value, decay with time. When an emotion’s intensity reaches a value near zero, this emotion is removed from the active emotions set.

Finally, for each generated emotion a memory trace is created, apart from emotions caused by a retrieval event. Created memory traces will have their event set to the emotion’s event parameter. The memory trace’s emotion parameter will be set to a copy of the emotion so that when the emotion’s intensity changes, it will not change in the memory trace. Lastly, the time stamp is defined as the current simulation time.

### 4.5 Application

As the Behavior module is highly dependent on the application into which the agent architecture was integrated, we will describe them together. The application consisted of a game in which the player controls an avatar (*meemo captain*) and through it can issue commands to several non-player characters (*meemo minions*). The objective is to guide the meemo minions in each level to reach an exit point. The avatar and meemo minions should not be hurt in the level.

The meemo captain’s behavior is defined by the architecture presented and by player commands. The meemo minion’s behavior is mainly defined by the architecture described. This behavior includes: expressing facial expressions corresponding to the most intense emotion felt; color saturation variation based on mood; presentation of a thought balloon when the agent’s displayed emotion was caused by a retrieval event; and path choice avoiding locations where negative events have occurred and favoring paths where positive events have occurred.

## 5. EVALUATION

We used the described application to get some insight into our main hypothesis: *Autonomous agents with episodic memory retrieval of emotionally relevant events, will be perceived as more believable, than similar agents without it.*

### 5.1 Methodology

We performed a non-interactive experiment (due to timeline and resource constraints). The group of participants (a total of 96) were mainly adults (95% having ages between 14 and 48). Furthermore, the group had a relatively balanced gender distribution: 51% male and 49% female.

Participants were exposed to a simple story in which the character’s behavior was initially driven by our architecture. Two agents are shown walking in a tunnel (meemo 1 and meemo 2) with neutral expressions. One of them (meemo 2) falls in a trap and dies, with the other one reacting by showing a *sadness* expression (see Figure4). This expression was caused by a reaction rule with negative values of *desirability-for-self* and *desirability-for-other* that matched the event sensed.

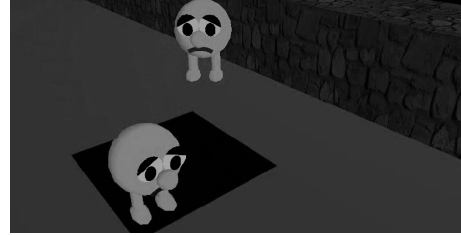


Figure 4: Evaluation Story - Part One

Afterwards the participant is explained that some time has passed, and sees a video showing meemo 1 going by the same tunnel and passing close by the trap, that is now easily avoidable. The character initially presents a *neutral* expression when entering the tunnel. The expression after passing the trap depended on the test condition.

The experiment had three test conditions: *retrieval, no retrieval* and *random expression*. In *retrieval*, the behavior of meemo 1 was driven by our agent architecture. When returning to the tunnel meemo 1 reacts emotionally, displaying a *sadness* expression. In *no retrieval*, the behavior of meemo 1 was simulated as if it was driven by an architecture with reactive appraisal but without episodic memory retrieval. Consequently, when returning to the tunnel, meemo 1 does not have any emotional reaction. Lastly, in *random expression* the meemo’s behavior is simulated as if it was driven by an agent architecture with reactive appraisal, without episodic memory retrieval, but with random expression of emotions. When the agent returns to the tunnel it displays a *happiness* facial expression. This outcome is only one of many that could possibly be generated by the architecture: the random generated emotional reaction needed not be in the tunnel; and the emotion expressed could be different. However, this architecture could only be truly tested with a longer exposure of participants to the agents’ behavior.

In an effort to do an objective analysis of believability, we indirectly measured it through believability features. Believability features are the participants’ perception of elements that are potential enhancers for believability. Among these believability features there were: *behavior coherence*, for in Ortony’s definition of believability [12] coherence is a crucial element; *change with experience* is one of Loyall’s requirements for believability [9]; *awareness*, that can be mapped to situated liveliness in [8]; and *behavior understandability*, for in Ortony’s definition [12], it is implicit that participants must be able to create a model of an agent’s behavior motivations. It is our belief that increased perception of these features translates into a greater sense of believability. Additionally, we also analyzed how participants graded meemos’ likability.

### 5.2 Results

When analyzing the values of behavior coherence, change with experience, awareness and behavior understandability we realized they were significantly higher ( $p < 0.025$ ) for test condition *retrieval* than for test condition *no retrieval*. On the whole results indicate that test condition *retrieval* was perceived as more believable than test condition *no retrieval*. This conclusion is consistent with our hypothesis.

Turning to the comparative analysis with test condition *random expression*, for change with experience, awareness and behavior understandability, the test condition *retrieval* did not present significantly higher values. We believe that one of the main contributing factors for this was the scenario's description not being very detailed thus allowing a wide range of interpretations.

On the other hand, test condition *retrieval* presented significantly higher values ( $p < 0.025$ ) of behavior coherence than test condition *random expression* (box plots for behavior coherence are presented in Figure 5). Being an important factor for an enhanced sense of believability, these results are also consistent with our hypothesis. Nonetheless participants would only get a clearer sense of agents' coherence after being exposed to several similar situations.

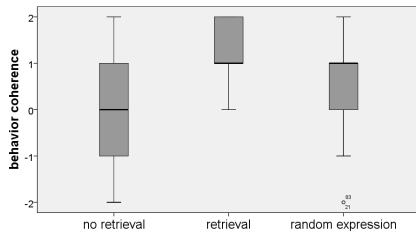


Figure 5: Box plots for behavior coherence

Additionally, we identified that the likability values were significantly lower for test condition *random expression*. Furthermore, some participants found meemo 1, in this test condition, to be “mean” or even “sadistic”. All these perceptions conflict with the meemo’s main design decision in the scenario: a reaction rule implying agreeableness. Finally, when analyzing if participants identified meemos’ emotions we achieved recognition rates between 74% and 97%.

## 6. CONCLUSIONS

Summing up, we have designed a model of episodic memory retrieval for believable agents inspired in human memory research. We implemented it, integrating it in a video-game application, and evaluated its impact in perceived believability. Results are coherent with the hypothesis that agents modeled by our architecture are perceived as more believable than agents modeled in similar architectures without episodic retrieval. Nonetheless, to analyze this hypothesis properly, further testing needs to be performed. In particular with a longer scenario in which agents are faced with a wider range of emotionally relevant episodes. To conclude, we believe our work represents a small step, yet relevant, towards modeling memory retrieval in agents and analyzing its impact on agent believability.

## 7. ACKNOWLEDGMENTS

This work is partially supported by the European Community (EC) and is currently funded by the EU FP7 ICT-

215554 project LIREC (Living with Robots and Interactive Companions), and FCT (INESC-ID multiannual funding) through the PIDDAC Program funds. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

## 8. REFERENCES

- [1] C. R. Barclay. *Schematization of autobiographical memory*, chapter 6, pages 82–99. Press, Cambridge University, 1988.
- [2] J. Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7):122–125, 1994. 176803.
- [3] C. Brom, K. Pešková, and J. Lukavský. What does your actor remember? towards characters with a full episodic memory. In *ICVS’07: Proceedings of the 4th international conference on Virtual storytelling*, pages 89–101, Berlin, Heidelberg, 2007. Springer-Verlag.
- [4] A. M. Colman. *Dictionary of Psychology*. Oxford University Press, third edition, 2009.
- [5] J. Dias, W. Ho, T. Vogt, N. Beeckman, A. Paiva, and E. André. I know what i did last summer: Autobiographic memory in synthetic characters. In *Affective Computing and Intelligent Interaction*, pages 606–617, 2007.
- [6] W. C. Ho, K. Dautenhahn, and C. L. Nehaniv. Computational memory architectures for autobiographic agents interacting in a complex virtual environment: a working model. *Connection Science*, 20(1):21–65, 2008.
- [7] W. C. Ho and S. Watson. *Intelligent Virtual Agents*, chapter Autobiographic Knowledge for Believable Virtual Characters, pages 383–394. Springer Berlin / Heidelberg, 2006.
- [8] J. C. Lester and B. A. Stone. Increasing believability in animated pedagogical agents, 1997.
- [9] A. Loyall. *Believable Agents: Building Interactive Personalities*. PhD thesis, Carnegie Mellon University, 1997.
- [10] S. C. Marsella and J. Gratch. Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.
- [11] A. R. Mayes and N. Roberts. Theories of episodic memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1413):1395–1408, 2001.
- [12] A. Ortony. *Emotions in Humans and Artifacts*, chapter On making believable emotional agents believable. MIT Press, 2003.
- [13] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Published by Cambridge University Press, 1990.
- [14] E. Phelps and T. Sharot. How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, 2008.
- [15] A. Rollings and E. Adams. *Andrew Rollings and Ernest Adams on Game Design*, chapter Character Development. New Riders Games, 2003. 1213088.
- [16] M. D. Rugg and E. L. Wilding. Retrieval processing and episodic memory. *Trends in Cognitive Sciences*, 4(3):108–115, 2000.



- [17] F. Thomas and O. Johnston. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, 1981.
- [18] E. Tulving. Episodic memory: From mind to brain. *Annual Review of Psychology*, 53:1–25, 2002.
- [19] E. Tulving, M. E. Voi, D. A. Routh, and E. Loftus. Ecphoric processes in episodic memory [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences (1934-1990)*, 302(1110):361–371, 1983.