Between Downward Spirals and Habituation: Emotion Intensity in Virtual Agents' Memory Retrieval

Paulo F. Gomes and **Ana Paiva** and **Carlos Martinho**[1]

Abstract. In the present article a model for memory retrieval of personal experiences for virtual agents is presented. It builds upon previous work and focuses on the effect memory retrieval can have on the agent's emotional state. Memory retrieval is defined as an emotional re-appraisal of past experiences. The variation of intensity of such re-experience is explored by modeling two phenomena: *downward spirals* and *habituation*. Downward spirals consist of situations in which recurrent retrieval of negatively charged past memories increases the recall intensity of such memories in depressed individuals. Habituation is the reduction of recall intensity caused by re-experiencing past experiences in a context perceived by the individual as safe.

1 INTRODUCTION

Recollection of personal experiences sometimes shoots oneself through a trip of sensations. These experiences can be moments of professional achievement, amorous conquest, peer conflict, a friend departing, etc. We travel to a past time, reliving those episodes [18], with previously elicited emotions coming to mind [12]. Daniel Schacter describes how troublesome this voyage can sometimes be:

"*We've all endured difficult experiences - the death of a loved one, rejection by a lover, failure at work - that pain us mightily in the days and weeks after they occur. In the immediate aftermath, we may find ourselves reliving the painful incident to the point of distraction,* ..."[17]

Should we not expect believable agents to have some degree of this "emotional memory retrieval"? How coherent would seem the behavior of an agent, that when returning for the first time to the place where it was first kissed, displays no emotional reaction? To address this issue, in previous work we defined and implemented a model in which an agent emotionally re-appraises past events when retrieving them [6][7].

However, the evaluation of the model seemed to indicate that because an agent tended to always react emotionally in the same way when retrieving a specific past memory, in the long-term this might render its behavior as excessively repetitive and predictable [6]. In the current document we propose a mechanism to add some variability to the model, by changing the experience's intensity between different retrieval situations. Furthermore, this mechanism will be inspired in two phenomena of human memory. Such a model can ultimately contribute to the believability of virtual agents by supporting behavior coherence while maintaining variability in the behavior [15].

[1] INESC-ID and Instituto Superior Técnico, Portugal, emails: pgomes@gaips.inesc-id.pt and ana.paiva@inesc-id.pt and carlos.martinho@ist.utl.pt

2 RELATED WORK

Many researchers have already defined agent architectures in which both past memories and emotions are present. The cognitive architecture Soar [11] is an example, yet in it there is no integration between the two. In [8] a virtual agent architecture framework is proposed in which past events are used to select strategies to deal with similar current events, and the emotional state is an indicator of the discrepancy between a desired state and the perceived state. However, emotion is not part of the memory retrieval process.

In FAtiMA [4] agents emotionally appraise events, store a personal story of these events, and can textually reconstruct it. Also oriented towards enabling reconstruction of personal stories, the agent memory system described in [3] takes into account memories' emotional charge. Memories concerning events that were initially perceived as more emotionally relevant take longer to forget. In opposition, in the cognitive architecture presented in [5], memory episodes linked with emotions of higher intensity have a higher probability of being retrieved. Nevertheless, in none of these architectures are past events re-appraised emotionally.

On the other hand, in the SALT & PEPPER agent architecture [2] emotion is an integral part of the retrieval process. The authors define emotions as performance evaluators and attention shift warnings. When an emotion is generated it is matched against the header of all nodes in a memory network. Nodes have different activation levels and only the matching node with highest activation level is retrieved to working memory. This retrieval causes the node's activation level to increase, activation which in turn is spread to neighboring nodes, similarly to activation spread in ACT-R's declarative memory [1]. Although it would be possible to use SALT & PEPPER to support change in the emotional state caused by retrieval of personal experiences, few clues are given on exactly how to discriminate between past and present in their appraisal.

In [10] the authors describe an agent architecture for an interactive virtual character (Eva) for which past interactions indirectly influence the agent's emotional state. The relationship between the agent and a specific user, defined in a two-dimensional space of dominance and friendliness, depends on emotions felt by the agent in previous interactions. For instance, gratitude will increase perceived user friendliness and decrease perceived dominance towards the user. Moreover, the relationship values change the agent's mood when interacting with the user, and in turn the mood affects the intensity of emotions. Although, memories indirectly influence the emotional state, the system is unable to model retrieval of past events not entailing a relationship.

All in all, supporting the inter-connection between emotion and memory has yet to be fully investigated, particularly in what concerns the retrieval process.

3 MEMORY RETRIEVAL MODEL

In previous work [7] a model for memory retrieval of personal experiences was defined, and implemented, drawing inspiration from Tulving's conceptualization of the process [19]. The model is divided in two main stages: *location ecphory* and *recollective experience*. In location ecphory, memories connected with the agent's current location are selected (details in [6]). The recollective experience consists of re-appraising the selected memories' associated past events according to current motives[2]. This model of memory retrieval was integrated in an agent architecture schematically represented in Figure 1.

As can be noticed, the recollective experience is essentially an appraisal process. This process follows closely FAtiMA's reactive appraisal [4] and the emotional related concepts are inspired in the OCC theory of emotions [15]. For instance, an *emotion* is defined as a valenced evaluation of an event and contains the following elements:

- *type* of the emotion according to the OCC model [15] (e.g. pity).
- *intensity* specifies the emotion's current intensity (non-negative scalar value). It decays to zero with time. When the intensity reaches a close to zero value the emotion is removed from the emotional state.
- *valence* specifies the emotion's value (positive or negative). The valence is directly dependent on the emotion type. For instance, joy emotions are positively valenced and pity emotions are negatively valenced.

Moreover, an *emotional state* is defined by the following elements:

- *active emotions* contains the set of emotions the agent is currently feeling. The *Behavior* module of the agent architecture (Figure 1) is responsible for making the agent display a facial expression corresponding to the emotion with highest intensity.
- *mood* is a bounded scalar value that represents the agent's recent overall emotional state valence. The evocation of negatively valenced emotions decreases the mood, while the evocation of positively valenced emotions increases it. These changes are proportional to the intensity of the evoked emotions. Additionally, mood decays to a neutral value with time.

The emotional state is affected by the appraisal of past events, during the recollective experience stage of memory retrieval, as well as present perceived events (regular appraisal). Events have a frame-like representation in which an event can have a sub-event. For example, a past event of witnessing another agent falling in a hole placed at bi-dimensional coordinates (200,300) could have the following representation: [Event type: retrieval sub-event:[Event type: witness sub-event:[Event type: fallHole location:(200,300)]]]. A past event is characterized by having type *retrieval*, and its sub-event parameter value is the event associated with the retrieved memory selected in the ecphory stage. Past events are stored in memory as memory traces. A memory trace is defined by the following elements:

- *event* is a representation of the event which the memory is about.
- *emotion* specifies the emotion caused by the appraisal of the event. If an event's appraisal generates more than one emotion, one memory trace is created for each one. In the remainder of this paper, memory traces with a negative emotion will be referred to as *negatively valenced*, and memory traces with a positive emotion will be referred to as *positively valenced*.
- *time stamp* is a meta-field indicating when the event started or when the memory trace was retrieved for the last time.

[2] The term motive is used as an abstraction over goals or desires.

In the current system only if an event's appraisal causes a change in the emotional state is the event stored. This approach tries to simulate the effect emotion content has on promoting attention focus during encoding [16] and on enhancing elaborative rehearsal. Instead of enforcing a forgetting mechanism such the one in [3], the events are filtered during storage. However, we have yet to empirically compare these two methods.

When a memory trace is selected by ecphory, a past event is created and fed to the appraisal system for recollective experience. This past event has type retrieval and has as its sub-event the event parameter of the memory trace. In the model, a past event appraisal consists of appraising its sub-event. However, appraisal may differ from the original one regarding the following perspectives:

- The mood influences the intensity of generated emotions: a positive mood enhances positively valenced emotions' intensity, and reduces the intensity of negatively valenced emotions. As mood may change, so can the intensity of generated emotions. Furthermore, if the re-calculated intensity is small enough, the emotion will not even change the emotional state. The mood influences appraisal by reflecting the recent past emotional experience.
- The agent's motives may change, and hence the appraisal and consequent generated emotions may differ from the originally created.

Finally, guided by the assumption that re-living a past experience will typically be less intense as the original experience [12], the final intensity of a generated emotion due to the appraisal of a past event is reduced by a parameterizable positive factor smaller than one (*memory retrieval intensity bias*):

$$intensity = intensity \times memory\ retrieval\ intensity\ bias$$

4 DOWNWARD SPIRALS AND HABITUATION

It is debatable to assume that re-living a past experience will always be a less intense experience than actually perceiving the experience first-hand, as the just shown expression implies. The two presented factors of emotion intensity variability have limitations: due to time decay, the mood only encodes the agent's recent past emotional context, and not an overall process of dealing with the past experience; changing motives arbitrarily can harm an agent's believability by preventing an individual evaluating the agent's behavior from creating a mental model of these motives [14]. Therefore, we propose adding to the existing model a mechanism for tuning emotion intensity variability during memory retrieval that takes into account the agent's overall experience with a specific memory.

Daniel Schacter discussed how retrieval of a certain memory can influence future recall of that same memory [17]. He states that continual reminding of a personal memory can sometimes strengthen recall. Moreover, that depressed patients better encode negative experiences and have a greater tendency for a phenomenon he names as memory *persistence*. Persistence of memories consists on an individual recurrently retrieving a memory when he, or she, does not wish to do so. An accountant not being able to concentrate on doing a company's tax declaration due to the constant retrieval of a car accident in which he was hurt, is an example of memory persistence.

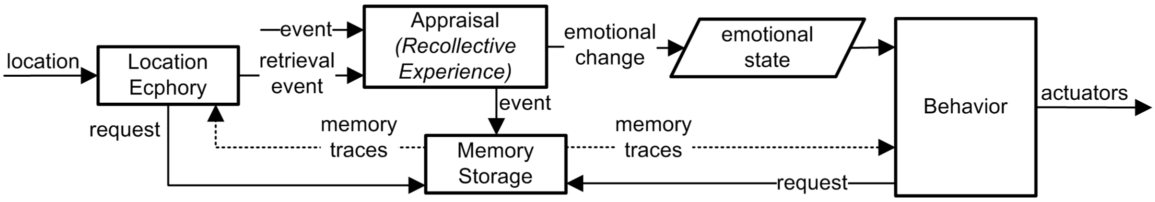

Figure 1. Agent Architecture

Depressed individuals are more prone to enter an emotional *downward spiral* connected with this phenomenon: they have a greater tendency to retrieve negative experiences, which can negatively influence their mood, consequently increasing their susceptibility to encode more negatives events (that can the experience of retrieval itself), that in turn will have a greater tendency to be retrieved. Schacter proposes that this loop may internally be enforced by an integration of negative past experiences into the self-schema (the individual's conception of itself) in which the former is linked with generic negative concepts such as "I do not do anything right" or "bad things always happen to me".

Despite their possibly devastating effects, Schacter argues that downward spirals caused by memory persistence may not be unavoidable. Telling others troubling experiences has been shown to have positive effects. Furthermore, re-experiencing a past traumatic experience in a safe context may result in a better integration with the self-schema and a less intense recollection further on. Abstractly, we can interpret this process as an *habituation*: reduced physiological response to repeated stimuli.

5 DYNAMIC EMOTION INTENSITY IN MEMORY RETRIEVAL

We propose modifying the memory retrieval model presented in Section 3 to take into account the phenomena of downward spiral and habituation. Each memory trace would have its own retrieval intensity bias value, that we rename as *persistence*, and this value could change when the memory trace was retrieved: if habituation took place it would decrease; if the downward spiral phenomenon took place it would increase; if neither phenomena were active the value would remain the same. Consequently, we need to define the conditions in which habituation and downward spirals take place.

Schacter presents depression as being a factor for an increased downward spiral effect [17]. Although depressed individuals are more prone to downward spirals, that does not rule out that the phenomenon, at least to some degree, can take place in non-depressed individuals. In a first approach, we propose mapping a state leaning towards potential depression in humans, to the agent having an extremely low mood value. Therefore, we propose that for the downward spiral phenomenon to take place, the mood must be lower than a *downward spiral threshold*. For instance, we could consider cases in which the mood is in the lower quarter of the mood range. Note that *downward spiral threshold* should be lower than the defined neutral value (see Section 3).

Turning to habituation, Schacter indicates the feeling of safety as promoting this phenomenon [17]. We propose mapping the feeling of safety to the absence in the emotional state of an emotion of type "fear" according to the OCC model [15]. Hence, for habituation to take place, there can not be an emotion of type "fear" in the emotional state.

Additionally, as habituation and downward spirals have opposite effects on subsequent recollective experiences of a memory trace, decreasing the emotion intensity and increasing it, we propose that they should be mutually exclusive. Therefore, we define that downward spirals will only take place if there is an emotion of type "fear" in the emotional state, and that habituation will only take place if the mood is higher than the downward spiral threshold. Summarizing the phenomena conditions:

- **downward spiral**: mood is lower than downward spiral threshold, and an emotion of type "fear" is present in the emotional state.
- **habituation**: mood is higher than downward spiral threshold, and no emotion of type "fear" is present in the emotional state.

Finally, considering the change in persistence (formerly named as memory retrieval bias), we propose that when a memory trace is retrieved in the downward spiral's conditions, the persistence is increased by the parameter value *persistence increase*. However, if with this increase the persistence would be greater than a parameter *max persistence*, then it would be reset to max persistence[3]. Analogously, when the habituation conditions are verified, the persistence could decrease to a minimum of *min persistence*. As the emotion intensity is a non-negative scalar, and the recalculated intensity is $intensity \times persistence$, min persistence must not be negative.

6 APPLICATION

The variation of the memory retrieval's emotion intensity will only positively affect believability if it is reflected on the agent's Behavior. The model presented in Section 3 has been integrated into a game prototype ("Meemos' Rescue") in which that happens [6]. In "Meemos' Rescue" the player controls a character (meemo captain) and through it can issue commands to several non-player characters (meemo minions). The objective is to lead the meemo minions to an exit point. The meemo minion's expressive behavior greatly depends on the architecture of Figure 1, and we will refer to them as agents for the remainder of this section.

An agent's mood is graphically represented by its color saturation. The lower the saturation of a color, the closer it will be to a gray tone. The word "gray" can be used to classify a mood [9] describing it as negative. In fact, gray tones are sometimes associated with a negative state of mind. With this motivation, when the mood is below its neutral value, the lower it is, the lower its color saturation is. Considering that the saturation percentage varies between 0 (minimum saturation)

[3] The value 1 is a potentially interesting candidate for the max persistence: on one hand the retrieval of past experiences will typically be less intense than the actual experience; on the other, if the mood is much lower than in the original situation, the resulting emotion can in fact be more intense than the original one.

and 1 (normal saturation), and that negative moods belong to the interval $[minimum\ mood; neutral\ mood]$ the saturation percentage is calculated by the following expression:

$$saturation\ percentage = \frac{mood - minimum\ mood}{neutral\ mood - minimum\ mood}$$

Mood is influenced by the evocation of emotions, as mentioned in Section 3. Consequently, the retrieval of negatively valenced memory traces indirectly decreases the agent's color saturation. Moreover, the decrease in color saturation will be proportional to the emotion intensity of the memory retrieval.

Still considering emotion expression, agents express the most intense emotion through a facial expression (see Figure 2). The current implementation supports four expressions: *neutral* (that acts as a baseline), *sadness*, *happiness* and *anger*. The choice of the three last expressions was motivated by the fact that they are part of the group of six universally recognized facial expressions (anger, disgust, fear, happiness, sadness and surprise) [13].

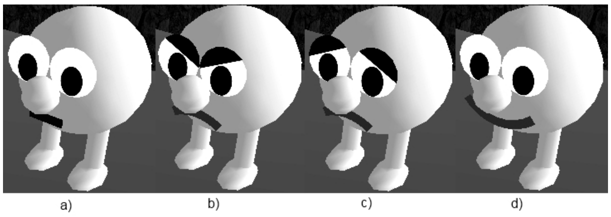

Figure 2. Meemo's face expressions: a) neutral b) anger c) sadness (Distress/Pity) d) happiness (Joy/HappyFor)

We mapped emotion types to the available emotion expressions. Note that each emotion type from OCC [15] represents a family of emotions. The emotion type "joy", for instance, represents emotional states such as happy, glad, delighted, pleased, etc. On the other hand, "distress" represents emotional states such as sad, unhappy, feeling bad, displeased, dissatisfied, etc. With this in mind, we mapped the emotion type "joy" to the *happiness* expression and the emotion type "distress" to the *sadness* expression. If there are no emotions in the active emotions, the *neutral* expression is displayed (Figure 2a). Unfortunately, emotions of type "angry" are not supported by our current model implementation, hence *angry* expression is never selected.

Turning to other supported emotion types, "happy-for" represents emotional states happy-for, delighted-for, pleased-for, etc. One can notice that these emotional states are quite similar to the ones presented for "joy". Thus, "happy-for" was also mapped to the *happiness* expression. "Pity" represents emotional states compassion, sympathy, sad-for, sorry-for, etc. It has been claimed that emotions such as compassion and sympathy have a different facial display pattern than distress [13]. However, this pattern seems to include oblique eyebrows, which are part of the *sadness* expression. Consequently, "pity" was also mapped to *sadness* expression. Lastly, the remaining emotion types supported by the implementation do not have a mapped facial expression.

In addition, if the agent's facial expression changes due to the retrieval of a past event, a thought balloon is presented (see Figure 3) for short period of time (parameterizable in a configuration file). An image is displayed on the thought balloon representing the remembered event. In Figure 3 the retrieved event was witnessing another agent falling into a trapdoor.

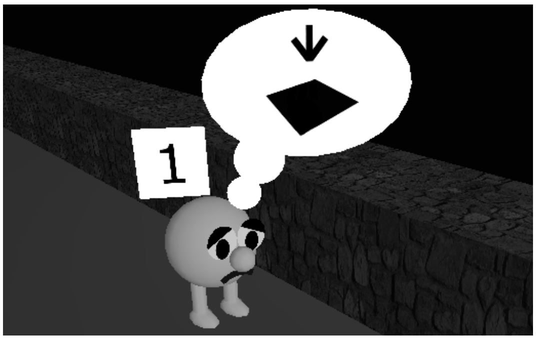

Figure 3. Meemo's thought balloon

Finally, the agent's behavior also directly considers the memories. The path planning tends to avoid locations where negative events have occurred and favoring paths where positive events have occurred [6].

7 EVALUATION

We believe that such an application can be used in a user study to evaluate how variating the emotion intensity of the recollective experience, as described in Section 5, can improve perceived believability. We propose a scenario in which a character negatively appraises an event taking place at a certain location, and returns to that same location several times, always remembering the past event (recall 1, 2 and 3). We can separate the evaluation in two sub-scenarios, one to evaluate the habituation effect (storyboard in Figure 4), another to evaluate the downward spiral effect (storyboards in Figure 5).

In the former, there will be two test conditions: *with habituation* (with H) and *without habituation* (without H). In test condition *with habituation*, when returning to the mentioned location, and remembering the past event (symbolized by the appearance of a though balloon with an iconic representation of the event), the character will progressively express a less intense expression. In test condition *without habituation* the reaction to the memory retrieval will always be the same.

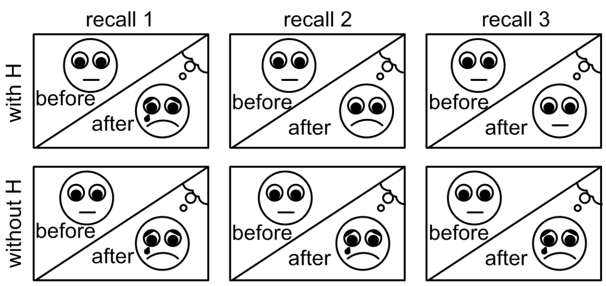

Figure 4. Habituation evaluation storyboard

When considering the effect of downward spirals, there will also be two test conditions: *with downward spiral* (with DS) and *without downward spiral* (without DS). In both test conditions, when returning to the mentioned location for the second time, the character displays a fearful expression. In test condition *with downward spiral* the emotional reaction of the character will progressively be more intense. In test condition *without downward spiral*, the reaction will always have the same intensity.

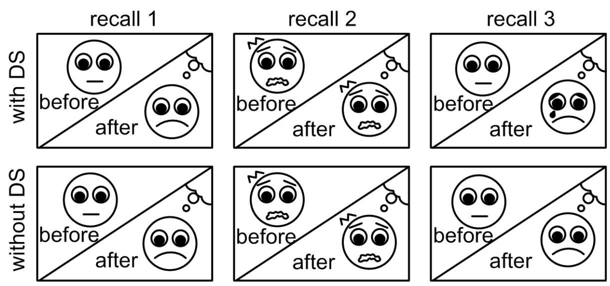

Figure 5. Downward spiral evaluation storyboard

As proposed in [14], believable character's behavior should be perceived as coherent, but at the same time not be perceived as excessively predictable. We propose that test participants, after being exposed to a test condition, would have to rate perceived coherence and predictability similarly to how it was performed in [6][7]. We expect that: test condition *with habituation* will present lower scores for predictability, and similar scores for coherence, when compared to *without habituation*; test condition *with downward spiral* will present lower scores for predictability, and similar scores for coherence, when compared to *without downward spiral*.

8 CONCLUDING REMARKS

In the present article a model for memory retrieval of personal experiences was presented. It builds upon previous work and focuses on the effect memory retrieval can have on the emotional state. Memory retrieval is defined as an emotional re-appraisal of past experiences. The variation of intensity of such experience is explored by modeling the phenomena of downward spirals and habituation. The modifications needed to take into account these phenomena have yet to be implemented and evaluated.

Although the modifications may account for a possibly greater variation of emotion intensity between retrievals, they unfortunately introduce four new parameters to the model: persistence increase, max persistence, persistence decrease and min persistence. Furthermore, it can be argued that the persistence increase should depend on the intensity of the active fear emotion or on the mood value. A potential approach for setting these parameters, as well as the initial persistence value, would be to map the differences in retrieval results between normal individuals and depressed individuals, to the differences in persistence between agents in conditions of the downward spiral phenomenon and agents not in these conditions.

If the "Meemos' Rescue" application is to be used to create the evaluation scenario, support for expressive behavior when the agent is feeling "fear" needs to be added (e.g. having the character's body tremble). It would also be an advantage to have different facial expressions for high and low intensity emotions.

Finally, as the model becomes increasingly more complex, the need for mechanisms such as forgetting, activation spreading, temporal organization between memory traces, and network organization of memory traces, also increases. Future work will probably entail integrating it into another memory system that already supports some of these mechanisms.

ACKNOWLEDGEMENTS

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. I would like to thank the anonymous reviewers for all the comments and suggestions.

REFERENCES

[1] John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin, 'An integrated theory of the mind', *Psychological Review*, **111**(4), 1036–1060, (2004).
[2] Luís M. Botelho and Helder Coelho, 'Machinery for artificial emotions', *Cybernetics and Systems*, **32**(5), 465–506, (2001).
[3] Cyril Brom, Klára Pešková, and Jiří Lukavsky, 'What does your actor remember? towards characters with a full episodic memory', in *ICVS'07: Proceedings of the 4th international conference on Virtual storytelling*, pp. 89–101, Berlin, Heidelberg, (2007). Springer–Verlag.
[4] João Dias, Wan C. Ho, Thurid Vogt, Nathalie Beeckman, Ana Paiva, and Elisabeth André, 'I know what i did last summer: Autobiographic memory in synthetic characters', in *Affective Computing and Intelligent Interaction*, pp. 606–617, (2007).
[5] Will Dodd and Ridelto Gutierrez, 'The role of episodic memory and emotion in a cognitive robot', in *IEEE International Workshop on Robot and Human Interactive Communication*, pp. 692–697, (2005).
[6] Paulo F. Gomes, *MEEMOs: Believable Agents with Episodic Memory Retrieval*, Master's thesis, Instituto Superior Técnico - Universidade Técnica de Lisboa, 2010.
[7] Paulo F. Gomes, Carlos Martinho, and Ana Paiva, 'I've been here before! location and appraisal in memory', in *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, (2011). To appear.
[8] Wan C. Ho and Scott Watson, *Intelligent Virtual Agents*, chapter Autobiographic Knowledge for Believable Virtual Characters, 383–394, Springer Berlin / Heidelberg, 2006.
[9] *The American Heritage Dictionary of the English Language*, Houghton Mifflin Company, fourth edn., 2009.
[10] Zerrin Kasap, Maher B. Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann, 'Making them remember - emotional virtual characters with memory', *IEEE Computer Graphics and Applications*, **29**(2), 20–29, (2009).
[11] John E. Laird, 'Extending the soar cognitive architecture', in *Artificial General Intelligence*, (2008).
[12] Andrew R. Mayes and Neil Roberts, 'Theories of episodic memory', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **356**(1413), 1395–1408, (2001).
[13] Keith Oatley, Dacher Keltner, and Jennifer M. Jenkins, *Understanding Emotions*, chapter Communication of Emotions, 83–114, Blackwell Publishing, 2006.
[14] Andrew Ortony, *Emotions in Humans and Artifacts*, chapter On making believable emotional agents believable, MIT Press, 2003.
[15] Andrew Ortony, Gerald L. Clore, and Allan Collins, *The Cognitive Structure of Emotions*, Published by Cambridge University Press, 1990.
[16] Elizabeth Phelps and Tali Sharot, 'How (and why) emotion enhances the subjective sense of recollection', *Current Directions in Psychological Science*, (2008).
[17] Daniel L. Schacter, *How The Mind Forgets and Remembers*, Souvenir Press, 2001.
[18] Endel Tulving, 'Episodic memory: From mind to brain', *Annual Review of Psychology*, **53**, 1–25, (2002).
[19] Endel Tulving, Martin E. Le Voi, David A. Routh, and Elizabeth 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).