Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.
Artificial Empathic Memory: Enabling Media Technologies to Better Understand Subjective User Experience

Bernd Dudzik*
Delft University of Technology
Delft, South Holland, The Netherlands
B.J.W.Dudzik@tudelft.nl

Mark Neerincx
Delft University of Technology
Delft, South Holland, The Netherlands
M.A.Neerincx@tudelft.nl

Hayley Hung
Delft University of Technology
Delft, South Holland, The Netherlands
H.Hung@tudelft.nl

Joost Broekens
Delft University of Technology
Delft, South Holland, The Netherlands
D.J.Broekens@tudelft.nl

ABSTRACT

An essential part of being an individual is our personal history, in particular our episodic memories. Episodic memories revolve around events that took place in a person’s past and are typically defined by a time, place, emotional associations, and other contextual information. They form an important driver for our emotional and cognitive interpretation of what is currently happening. This includes interactions with media technologies.

However, current approaches for personalizing interactions with these technologies are neither aware of what episodic memories are triggered in users, nor of their emotional interpretations of those memories. We argue that this is a serious limitation, because it prevents applications from correctly estimating users’ experiences. In short, such technologies lack empathy.

In this position paper, we argue that media technologies need an Artificial Empathic Memory (AEM) of their users to address this issue. We propose a psychologically inspired architecture, examine the challenges to be solved, and highlight how existing research can become a starting point for overcoming them.

CCS CONCEPTS
• Human-centered computing → Ubiquitous and mobile computing theory, concepts and paradigms;

KEYWORDS
Personalization; Media-evoked Emotions; Affective Computing; User Modeling; Empathic Technology; Episodic Memory

1 INTRODUCTION

User-modeling describes approaches that enable interactive applications to adapt their behavior dynamically in response to the specific preferences and needs of their users, thereby creating more personalized experiences [34, 48]. Insights into users’ emotions are a particularly valuable source of information for applications about how interactions with them are being experienced [44]. For this reason, attempts are being made to provide media technologies with the ability to relate to peoples’ emotions (e.g. [41]). For example, applications attempt to anticipate the emotional qualities that interacting with specific media content evokes in users for the purpose of facilitating personalized recommendations (e.g. [3, 69, 81]).

However, making such estimations correctly is a challenging task. Individuals’ emotional experiences of events are highly subjective and express a dynamic relationship between them and their current situation [56]. As such, the emotional qualities of media experiences may vary significantly across people and in response to the specific context in which they take place [36].

One important contextual influence on how people experience their present situation are the personal memories that it brings to mind. Moments in which human beings re-experience specific events from within their personal history are known as Episodic Memories [21]. These recollections typically include a sense of the time and the place at which remembered events have occurred, as well as potentially vivid visual imagery [21]. These memories also may contain strong affective associations that have an influence on our present emotional interpretations [7], and emerge spontaneously in response to our current environment [8]. Interactions with media content, such as personally meaningful pieces of music, may be particularly affected by this emotional influence from the past: empirical research has identified that these can act as potent triggers for episodic memories about events from a listener’s past [47, 65]. Moreover, the emotional tone associated with the memories elicited in this way, has a strong influence on how listening to a piece feels [6, 82].

Such influences of a user’s memories on the subjective experience of interactions with media technology poses three interpretation challenges that need to be addressed by personalization technologies:
• Knowing the conditions that trigger episodic memories. We refer to this as the challenge of receptiveness, i.e., whether an individual is receptive to remembering episodes in the current situation;
• Knowing what episodic memories of the user are triggered by the current situation. We refer to this as the challenge of content, i.e., which memories are triggered by the current situation;
• Knowing the emotional interpretation of a remembered event and its impact on both the interpretation of the current situation and emotional state of the user. We refer to this as the challenge of appraisal, i.e., how does the current situation, including the triggered episodic memories, feel for the individual.

We summarize these challenges as the RECAP problem (REception, Content, and APpraisal). Approaches for personalization of media interactions that fail to address the RECAP problem are unable to provide reliable estimations of a user’s experiences. This is because they are unable to predict if memories impact the user’s experience of the current situation at all, what memories impact it, and how they impact it. In a very real sense, being recollection-unaware in such a fashion makes technologies lack empathy, since they cannot relate in any way to the influence that having access to a personal history exercises on their users’ feelings.

In this position paper, we argue for the need to create recollection-aware technologies that are capable of addressing the RECAP problem. Furthermore, we propose a generic architecture for the simulation of episodic memory processes in users through an Artificial Empathic Memory (AEM) as a means to do so. It provides application with the ability to reason about how episodic memories influence their users’ cognitive-affective interpretations of a specific situation, thereby enabling them to more thoroughly empathize with their subjective experiences.

2 THE RAMIFICATIONS OF RECOLLECTION-UNAWARE MEDIA TECHNOLOGY

In the following, we discuss challenges and limitations in interactions between users and multimedia applications that are unable to relate to the way that human beings dynamically experience their own past. We begin by discussing the implications of the RECAP problem on interactions with multimedia content in an example application domain where they are particularly salient, before describing broader consequences for personalizing users’ experiences with media technologies.

2.1 An Example: Challenges for Media-based Reminiscence Support Technologies

Reminiscing about the past is a complex activity that involves recollection, interpretation and often sharing of personally significant memories [11]. Reminiscing has been attributed with the fulfillment of numerous psycho-social functions for individuals engaging in it [10, 27]. Additionally, guided reminiscence activities are being successfully used as interventions to improve subjective well-being [17] and ameliorate depression [43]. Motivated by this, researchers have displayed an ongoing interest in designing technologies to support and encourage such activities (e.g. [5, 25, 49, 55]). A popular strategy consists of attempting to evoke recollections in users by presenting them with personally relevant media content [24, 58, 60, 76, 77]. Content shared on social media platforms is a particularly rich and accessible resource for these purposes (e.g. [66]), and recently has been exploited in large-scale applications created by commercial enterprises to spark recollections in users (e.g. [33, 71]).

Despite a persistent interest in the development of technologies that promote and support reminiscence activities through media content, relatively little is known about their effectiveness. Initial empirical findings indicate a potential for evoking meaningful and desirable recollections in users, but they also highlight how the RECAP problem hinders applications to do so successfully [25, 61, 66].

First, individuals are not always in a state of mind in which external stimuli trigger episodic memories. One factor that appears to contribute to their capacity to do so is individuals’ willingness and ability to think about their past [25]. Empirical findings indicate that the capacity of an attempt to evoke episodic memories depends on how absorbed individuals are in the activities that they are currently undertaking [61, 78]. This demonstrates the importance solving the challenge of receptiveness for this kind of applications.

Second, content provided as trigger may not be successful at inducing any episodic memories in individuals, even in cases where it objectively documents moments from within their personal past. In these situations, people may recognize that a relation to themselves exist (e.g. that it is them who are being depicted in a photograph), but have no recollection [25, 61]. Moreover, the same stimulus may bring to mind varying memories when encountered at different moments in time, or result in memories of multiple events. In short, it is unknown what memories are evoked by a particular trigger. This is an instance of the challenge of content.

Third, undergoing episodic memories of certain events may be experienced as undesirable by users of an application. For example, people have reported that applications evoke recollections of events from their life that they did not experience as significant or interesting enough to warrant their attention in the present [25]. In other studies, participants have even described negative emotional responses towards being reminded of certain events from their past [46, 50, 61]. Strikingly, existing applications of reminiscence technologies integrated in social media platforms, have led to recollections of unwanted memories, undermining their explicit goal of providing their users with joyful experiences [13, 38]. Solving the challenge of appraisal is therefore an important step towards predicting the user’s emotional interpretation of a memory evoked by a multimedia stimulus.

Overall, these findings illustrate the implications of failing to address the RECAP problem for media-based reminiscence support technology. Because existing approaches to these applications are recollection-unaware, they are very limited in their capacity to cause episodic memories that are aligned with the goals of applications and the desires of their users. To function reliably, these applications need to be capable of providing highly personalized experiences, i.e. they require recommendation of multimedia stimuli that are meaningful for an individual user, in light of his or
her personal past in a specific situation. This cannot be facilitated without an approach to user-modeling and context-awareness that addresses all aspects of the RECAP problem.

2.2 Broader Implications for Personalizing Interactions with Media Technology

The impact of not addressing the RECAP problem stretches beyond this specific application setting, however, and has broad implications for personalizing users’ interactions with media technologies.

For once, any form of recommender system can benefit from addressing the RECAP challenge, both w.r.t. to what suggestions they make, as well as in choosing the means for how they do it. Contemporary versions of such systems are location and time aware and can, for example, suggest dinner locations with the understanding that it is lunch time and where the nearest dining locations are. However, these suggestions are unable to take into account the local haunts of an ex-lover or suggest a restaurant because a meaningful family celebration for that individual has occurred there. In this case, the experience of the recommendation is highly influenced by the appraisal of the episodic memories associated with the dinner location. This appraisal may change over time and due to other factors than the person’s experience at those locations. Moreover, the form in which the recommendation is provided (e.g. involving a photo) may accidentally trigger memories that influence its experience significantly. Without a model to simulate the recollection of a memory and its appraisal in the present, it is practically impossible to predict the emotional experience associated with a particular recommendation.

Another example is human memory support through media-based reminders (see [42] for a recent overview). Addressing the RECAP problem is a core challenge that these systems face for the provision of functional and personalized support. They need to understand when individuals may be able to actually benefit from external reminders (Receptiveness), determine how to best elicit a specific memory with the modalities available to them (Content), and estimate the potential emotional consequences of reminding their users of a specific event (Appraisal). In a similar vein, the RECAP problem holds relevance for personalizing interactions with e-Learning applications and intelligent tutoring systems. Here, content provided to students may accidentally trigger task-irrelevant (potentially emotionally charged) thoughts about their past. These may negatively impact their efficiency of learning by preventing them to focus on their intended learning objectives. Being able to estimate whether certain material or activities are likely to result in a state of mind in which such distracting memories emerge more easily provides valuable information to applications for personalizing their interactions (see e.g. [9]).

This is just a small selection of applications that are impacted by the RECAP problem. Because episodic memories form such an important part of human cognitive-affective functioning, numerous other scenarios can be envisioned where users’ experiences of interactions with media technologies could be improved.

3 AN ARCHITECTURE FOR ARTIFICIAL EMPATHIC MEMORY

Human beings have an innate ability to estimate how other people think and feel in response to events in their environment [45]. An important part of this empathic understanding is the cognitive-affective reasoning by which a person simulates the mental states of others, based on prior (shared-) experiences and general knowledge [28]. This is often referred to as Theory of Mind. Empirical findings suggest that the more familiar one person is with another, the more likely they are to gain accurate insights into how that other person feels [70]. In essence, to achieve true personalization of their interactions, applications need to possess a rudimentary theory of mind of their users. This would enable them to simulate what a user is actually thinking and feeling. Building an accurate and usable artificial theory of mind is, however, a bridge too far in the context of reliable computational modeling.

In this paper we argue that an important subset of that can already help to address the RECAP problem for providing personalized experiences. To that end, we propose a computational architecture for an Artificial Empathic Memory (AEM). It provides applications with the ability to predict the user’s experiences of a situation (including the system’s actions) while taking into account the episodic memories that are so important for forming his/her personal interpretation of it.

We argue, that for each of the three RECAP challenges there is a suitable psychological theory that can form the foundation of a functional component to address it computationally. In the following, we provide an overview over each of these components. In particular, we outline how they interact with each other to detect the individual’s attentional engagement in a present activity, predict the associative strength existing between external stimuli and episodic memories, and finally predict their impact on the emotional experience of individuals. See also Fig. 1 for an overview.

- Flow Detection Component: The input of this module consists of features describing a user’s current activity and state, while the output is the degree of attentional engagement that is experienced. This value modifies the operation of the ephoric processing-module: a low degree of attentional engagement results in a low activation threshold for episodic memories, biasing the Ephoric Processing Component to propose candidates for recollection.

- Ephoric Processing Component: The input of this component is the current situation (state + activity). It extracts (a subset of) the user’s current situation as an Episode in a representation that allows associative strength to be calculated (e.g. a vector of features). Then it determines the associative strength between that encoding of the situation and all available episodes in the Episodic Memory Store. The outcome of this operation are one or more Episodic Memories. The Episodic Memory Store forms an important resource for this process. It is a database that contains a collection of information about personal events from a user’s past in the form of encoded situations that we refer to as Episodes.

- Emotional Appraisal Component: This module simulates a series of cognitive-affective processes that determine the emotional quality of experiencing an episodic memory. It
3.1 Flow Detection and the Challenge of Detecting Receptiveness

3.1.1 Psychological background. Findings of empirical studies investigating the emergence of episodic memories in everyday life have demonstrated that they have a tendency to occur in situations where a person’s attention is not fully immersed in an ongoing activity [78, 79]. In this section, we argue that Flow is a useful psychological concept to understand and model the degree to which a user’s current situation gives rise to such attentional engagement.

The concept of flow describes a state of mind in which a person is so absorbed in performing an activity that there is no room for other thoughts to emerge [57]. A requirement for flow experiences to emerge, is that an ongoing situation holds a balance between the challenges that it presents to a person and his or her perceived ability to cope with them [26]. Importantly, situations that are experienced as lacking in challenge result in states of cognitive under-stimulation, e.g. boredom [54]. Here, individuals’ attentional resources are no-longer fully invested in the activities they are undertaking, thus creating conditions that are more favorable for episodic memories to emerge.

In summary, Flow is a concept that is widely used and empirically well established across a broad variety of disciplines. It provides a suitable theoretical framework to characterize how individuals’ degree of attentional engagement varies under specific circumstances, which in turn modulates individuals’ tendency for episodic recollection.

3.1.2 Computational Approaches Towards Flow Detection. A large body of work on flow and engagement detection exists, within the domains of entertainment and education computing. For example, research on detecting tutoring engagement showed initial successes at discriminating between flow-relevant states of boredom, frustration and confusion in learners [30]. In the adaptive gaming domain, automatic detection of boredom and frustration was also shown to be feasible [19, 67]. In some cases, these attempts reached a reported accuracy of over 90 percent in post-hoc classification of engagement and frustration based on recorded visual and game-play features [67]. Further, in the field of Human Robot Interaction, initial investigations have shown the possibility to detect engagement based on task and interaction-related features [18], in essence replicating findings in the gaming and e-learning/tutoring domain. Finally, research in the field of interruptibility detection (e.g. [35]) strongly relates to detecting flow and engagement based on a user’s context as captured by ubiquitous sensing technology [75]. In essence, these different areas seem to converge on similar ideas, namely, that it is both important and computationally feasible to detect task engagement in users. With the right focus, we believe these techniques can be extended to engagement measures that are correlated with the emergence and intensity of episodic recollections.

3.1.3 Challenges in Flow Detection. Important research challenges remain. First, significant sensing abilities are needed to detect engagement in users, in particular when focusing on social signals. Pupil-dilation might be an interesting alley for future research, as it seems to correlate with, for example, high temporal resolution attention dynamics [80]. As such, it might be an easy-to-detect, unimodal option for the detection of flow. As eye-tracking can now be reliably done using machine learning on data coming from standard cameras embedded in mobile devices [23], it is to be expected that pupil dilation detection too becomes feasible in the near future. This opens up the possibility to detect attentional engagement in real time on standard customer devices.
Another challenge hinted at by flow theory is that activities may result in varying degrees of attentional engagement for different individuals, because these do not experience the same degree of challenge. Detection in these circumstances can probably be enhanced with personalized engagement models. Information about a user’s specific skills or interest in particular activities might help a computational model of flow-processing to become more accurate in detecting momentary engagement.

Finally, flow detection will also need to be taken into account when compiling data traces into digital records describing persons’ episodic stores. This is because the information sensed by technological monitoring may not be aligned with the deployment of attentional resources by a person in the same situation. It may therefore provide a description of the events in question that is strikingly different from that person’s memories. Research on modeling human attentional focus (e.g., [63]) holds potential for improvement of this circumstance, e.g. by enabling applications to construct a model of a situation that corresponds more closely to the user’s perception of it.

3.2 Ecphoric Processing and the Challenge of Predicting Episodic Content

3.2.1 Psychological background. The degree of association between a present situation and an instance in a person’s past plays an essential role in the emergence of episodic memories with a specific content in contemporary psychological models of human memory (e.g., [22, 72]). For example, it is understood that the potential of a present situation to cause an episodic memory of a specific past event (i.e. to act as a memory cue for it), is dependent on its similarity to the context under which that event was originally committed to memory [68, 73]. The greater this overlap, the more likely it is to come to mind. Consequently, the associative strength of external stimuli both influences whether something comes to mind, as well what something is. However, the nature of the associations linking external stimuli to past events can take numerous forms and exist at different levels of abstraction. They can range from purely perceptual similarities between cues and elements of a past episode to associations that exist solely at a conceptual level [52].

One way to conceptualize the process of how stimuli act as cues for recollections of specific events has been proposed by Tulving [74]: in an initial phase called ecphory, cue attributes are correlated with information stored in memory as traces. The outcome of this process describes the potential activation of each trace given its association with the current cue [72, 74]. This is followed by a conversion-stage, in which the degree of activation determines whether the information in a trace is recollected or not [72]. This model provides a simple theoretical framework to conceptualize the influence that a situation has on the occurrence and content of episodic memories.

3.2.2 Computational Modeling of Episodic Memory Processes. Research in the domain of artificial intelligence has produced several computational models of episodic memory that implement retrieval mechanisms akin to ecphoric processing (e.g., [15, 37, 59, 62]). A common approach is to represent both cues and traces as an array of features, and to calculate the associative strength between them using a form of distance metric. Overall, a variety of plausible models of ecphoric processing exist in the agent and cognitive modeling fields. This is of importance as it means that, when such models can be populated with actual experience-rich content from users, they can be a start to simulate their episodic memory processes. This can be combined with a data-driven approach where a model learns over time which associations are more likely to occur for a person by receiving explicit feedback from them.

3.2.3 Challenges in Ecphoric Memory Processing. Several challenges for a computational model of ecphoric processing are important to discuss here. First, in order to facilitate a useful simulation of the evocative potential of situations, it is necessary to develop a representation for them that captures their potential to act as memory cues. Developing such a representation is challenging, since it must capture attributes at different levels of abstraction, i.e. facilitate both perceptual and semantic associations.

Second, the detection of what attributes of a situation are relevant for the process of memory elicitation is a difficult and unsolved problem. The main challenge here is that a stimulus can only act as a cue in a situation if a user is actually perceiving it. So, either a system must be certain that he or she attends to it due to the context of its presentation (e.g., in the case of it taking a large amount of screen estate), or we need means to estimate the target of a user’s attention (e.g, through detecting users’ attentional focus via gaze-tracking, see [63]).

Finally, a crucial resource required in ecphoric processing is a collection of personal information that describes those past experiences that may potentially resurface as episodic memories. In our architecture, these form the records of the Episodic Memory Store. One particular challenge here is the comprehensiveness required from these records: to meaningfully contribute in overcoming the RECAP problem, the they need to cover enough ground about users’ lives to facilitate association with the events relevant for a given situation. A starting point for its construction can be the substantial research on the creation of lifelogs. It describes the collection and organization of large quantities of data describing a person’s experiences into a single comprehensive digital repository [39]. Common tasks for constructing such an archive include recording and fusing multimodal data traces into a single timeline, its automatic segmentation into a structure of distinct events [31], and its automatic semantic annotation through pattern recognition techniques [40].

While addressing policies for population and management of such an episodic store is beyond the scope of this article, we feel it is important to highlight the challenge of maximizing privacy in its construction (both of users themselves and the people they encounter in their lives). Additionally, this includes methods for providing users with control over what parts of the episodic store is available for personalization purposes. Both privacy in lifelogging [51], as well as management of long-term user models [4] are the subject of ongoing research.
3.3 Cognitive Appraisal Theory and the Challenge of Predicting Emotional Experience

3.3.1 Psychological background. A series of common evaluative judgments (e.g. novelty, goal-congruence, etc. [56]) have been identified to reliably accompany and discriminate between emotional experiences [64]. These judgments can be seen as partial mental representations of the emotional qualities of experiences [2]. The view of relating cognitive judgments of personal meaning to emotional responses is called Cognitive Appraisal Theory (CAT). Its central assumption is that an organism’s emotional responses express how much personal significance it assigns to the information it processes in a given situation w.r.t its utility for the fulfillment of its concerns [56]. While the descriptions of appraisal processing given in the literature often focus on evaluations of individuals’ immediate surroundings, CAT argues that emotional appraisal is a fundamental mode of cognitive-affective functioning. As such, it applies to any kind of experiential content: perceived, remembered and even imagined [14]. In summary, CAT provides a general theoretical lens for understanding and describing the emergence of emotional qualities in experiences based on information processing. Because of this, we see them as a promising approach to model the relationship between episodic memories and the emotional qualities of their experience for a person.

3.3.2 Computational Modeling of Emotional Appraisal. Numerous computational models have drawn on appraisal theories to enable virtual agents or robots to display plausible emotional reactions to events in their artificial environments or in interactions with users (see [53] for a comprehensive overview). In addition, artificial intelligence research has used appraisal theory to enable virtual agents to reason about the potential emotional reactions of human beings that they are interacting with (e.g. [12]). Despite the popularity of appraisal theories as inspiration for computational models of emotion elicitation, they have not seen wide usage in models of experiencing episodic memories. However, several existing computational memory models for intelligent agents include an abstract representation of their emotional state (e.g. [16]) or appraisal values [29] to describe the emotional experience that an actor has had in a previous event. This work shows that it is feasible to model the appraisal of events in a personal context. Although work on appraising the situation and memories of an actual person (rather than a virtual agent or robot) is scarce, the modeling technique can be similar.

3.3.3 Challenges in Cognitive Appraisal of Episodic Memories. With the exception of [37] there has been no research on computational modeling of how episodic memories are appraised upon recollection. This may be in part because there are some conceptual challenges to a straightforward application of established appraisal theories to episodic memories as stimulus events that form the target of appraisal. Especially challenging is the fact that episodic memories contain multiple aspects that can be appraised by individuals. On the one hand, there is the recollected information itself (which already has been appraised in the past during the original experience). On the other hand, there is information available that describes the current circumstances under which the event is recollected, such as its relevance for a person’s current motivations. How these different sources of affective information shape the outcome of a person’s emotional interpretation of a situation needs to be accounted for in a computational model of this process. For this reason it is important to investigate how common dimensions in appraisal theories can most meaningfully be applied in a computational model of episodic memories, as well as in how far such an application produces outcomes that are plausible and congruent with human experiences at the moment of recollection.

An additional challenge is the inference of a person’s current goals. CAT postulates that motivations play an essential role in appraisal processing, but these constructs cannot be directly observed in individuals. As such, research contributing to their automatic inference from individuals’ behavior has a tremendous potential for supporting the computational modeling of emotional appraisal processing in users. Existing technologies, such as data-driven and automatic driver intention recognition (for short term goals) [32] and explicit preference elicitation (long term values and preferences of people) [20], demonstrate that this is at least a feasible road to take. Furthermore, existing research on activity recognition [1] can be already used for coarse goal detection (going to work, going to bed, etc.). As such, there is quite some work showing that the inference of users’ goals and intentions at different time scales is at least a feasible enterprise, given sufficient sensor data.

4 SUMMARY AND CONCLUSION

Experiences from our past are a primary influence on how we understand our environment in the present, including during interactions with multimedia applications. Ignoring this influence results in recollection-unaware media technology that is oblivious to the RECAP problem for personalization. Ramifications become strikingly evident when looking at scenarios where the primary goal of applications hinges on their capacity to shape experiences through elicitation of episodic memories.

We have argued that providing media technologies with increased empathy for their human users requires enabling them to display awareness of when and how they dynamically experience their past in episodic memories. Our proposed architecture for an Artificial Empathic Memory forms a psychologically-grounded computational blueprint for providing applications with the means to do this. It comprises a series of processing components that jointly form a computational model of how externally triggered episodic remembering influences the emotions evoked during interactions with media technologies. Access to this information enables applications to adjust their behavior in meaningful ways, thereby facilitating truly personalized experiences.

Instantiating the individual components of such an AEM is a challenging task. However, it benefits from existing technological research in a variety of areas, such as the detection of attentional deployment from multimodal sensor data, computational cognitive modeling, and the development of lifelogging appliances. Given this, we feel that there is no fundamental technological hurdle for developing applications that better understand their users’ subjective experiences by accounting for the role of episodic recollections in them.