Modelling autobiographical memory loss across life span

Wang, Di; Tan, Ah-Hwee; Miao, Chunyan; Moustafa, Ahmed A.

2019

Wang, D., Tan, A.-H., Miao, C., & Moustafa, A. A. (2019). Modelling autobiographical memory loss across life span. The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 1368-1375. doi:10.1609/aaai.v33i01.33011368

https://hdl.handle.net/10356/139099

https://doi.org/10.1609/aaai.v33i01.33011368

Modelling Autobiographical Memory Loss Across Life Span

Di Wang[1], **Ah-Hwee Tan**[1,2], **Chunyan Miao**[1,2,3], **Ahmed A. Moustafa**[4,5]
[1]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly
[2]School of Computer Science and Engineering
[3]Alibaba-NTU Singapore Joint Research Institute
Nanyang Technological University, Singapore
[4]School of Social Sciences and Psychology, Western Sydney University, Sydney, Australia
[5]Department of Social Sciences, College of Arts and Sciences, Qatar University, Doha, Qatar
{wangdi,asahtan,ascymiao}@ntu.edu.sg, a.moustafa@westernsydney.edu.au

Abstract

Neurocomputational modelling of long-term memory is a core topic in computational cognitive neuroscience, which is essential towards self-regulating brain-like AI systems. In this paper, we study how people generally lose their memories and emulate various memory loss phenomena using a neurocomputational autobiographical memory model. Specifically, based on prior neurocognitive and neuropsychology studies, we identify three neural processes, namely overload, decay and inhibition, which lead to memory loss in memory formation, storage and retrieval, respectively. For model validation, we collect a memory dataset comprising more than one thousand life events and emulate the three key memory loss processes with model parameters learnt from memory recall behavioural patterns found in human subjects of different age groups. The emulation results show high correlation with human memory recall performance across their life span, even with another population not being used for learning. To the best of our knowledge, this paper is the first research work on quantitative evaluations of autobiographical memory loss using a neurocomputational model.

Introduction

In recent years, many governments and agencies have invested a record-high amount of resources to look deeper into human brain's functional dynamics. However, as of today, it is still difficult or impossible to quantitatively evaluate a wide range of brain dynamics at the neural network level. From the point of view of AI, neurocomputational models built upon neurocognitive and neuropsychology theories can provide insight into human behavioural processes in a rapid and quantitative manner. For example, according to Wang, Gauthier, and Cottrell (2016), "one advantage of computational models is that we can analyse them in ways we cannot analyse human participants to provide hypotheses as to the underlying mechanisms of an effect."

In this paper, we evaluate how people generally lose their memories by exploiting an established computational autobiographical memory model (Wang, Tan, and Miao 2016), named Autobiographical Memory-Adaptive Resonance Theory network (AM-ART). AM-ART is built upon the psychological basis presented by Conway and Pleydell-Pearce (2000), which has been widely accepted and supported by neural imaging evidence (Addis et al. 2012). Our prior work (Wang, Tan, and Miao 2016) focuses on memory retrievals using imperfect cues and "wandering in reminiscence", which refers to recalling a sequence of seemingly random but contextually connected memory across different episodes of life events. In that prior work, we assume that the memory formation and retrieval processes can always be performed perfectly, which would rarely be true in real-world scenarios. Moreover, due to the hardware constraints in agents or robots, discard of certain portion of the stored memory is necessary in most complex application domains. Therefore, with a totally different purpose, in this paper, we show AM-ART can accurately emulate various human memory loss phenomena.

Specifically, we employ three key processes in AM-ART to replicate the three widely studied memory loss phases, namely during memory formation, storage and retrieval (Jahn 2013), respectively. Moreover, we introduce three novel parameters to AM-ART to regulate the corresponding memory loss processes, namely overload as the likelihood of being affected by cognitive overload during formation (Daselaar et al. 2009), decay as the rate of long-term memory fading during storage (Rubin 1982), and inhibition as the likelihood of retrieval failure during retrieval (Storm and Levy 2012). Our approach of using a neural network with relevant control parameters to model memory loss aligns with cognitive experts' opinion that "the individual pattern of impaired memory functions correlates with parameters of structural or functional brain integrity" (Jahn 2013).

For performance evaluations, we collect an autobiographical memory dataset comprising more than one thousand life events from public domains. However, because this collected dataset does not span across one's entire life (e.g., from childhood to 70s), in order to conduct relevant experiments, we alter the event dates so that the collected life events are equally distributed across the life stages and the ratio among pleasant, neutral and unpleasant memories in each life stage conforms to the distribution reported by Berntsen and Rubin (2002). Moreover, it has been found that people of all ages tend to recall more pleasant memories rather than unpleasant ones, although the voluntarily non-recalled unpleasant memories are still retained. We model

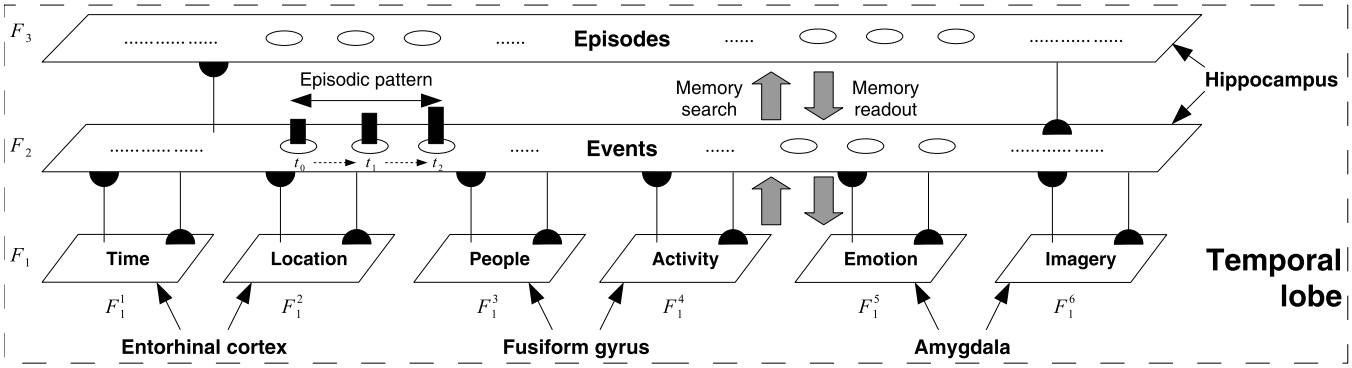


Figure 1: Network structure of AM-ART. All its input channels in F_1 and the F_2 and F_3 layers match specific brain regions.

this tendency based on the memory survey data reported by Rubin and Berntsen (2003). Subsequently, we perform model evaluations based on the memory recall data reported by Berntsen and Rubin (2002). Specifically, we learn the memory loss parameter values by emulating the memory recall performance of human subjects in different age groups and further use the learnt parameter values to predict the performance of human subjects in the subsequent life stage. The emulation results show high correlation, even with the memory recall performance of another population reported by Rubin and Schulkind (1997).

As such, we show that AM-ART can accurately capture the characteristics of human autobiographical memory loss. Therefore, we provide a useful tool to analyse various memory loss phenomena that may be difficult or impossible in human subjects. To the best of our knowledge, this paper is the first research work on quantitative evaluations of autobiographical memory loss using a neurocomputational model.

Related Work

For the same purpose of using a neurocomputational model to verify neurocognitive theories and perform quantitative evaluations, Wang, Gauthier, and Cottrell (2016) use PCA (principal component analysis) and MLP (multi-layer perceptron) with one hidden layer, wherein different number of hidden neurons are used to represent the corresponding level of the human participants' pattern recognition ability. Their model supports the "experience moderation effect" observed by Gauthier et al. (2014). In this paper, we use AM-ART as the neurocomputational model to replicate human memory loss phenomena in different age groups.

Many well-established cognitive models, such as Soar (Laird 2012), ACT-R (Anderson et al. 2004) and Icarus (Langley 2006), employ functionally specific memory modules. Moreover, few such cognitive models further investigate the dynamics of long-term memory forgetting, e.g., Derbinsky and Laird (2013) heuristically define memory decay mechanisms in Soar. Nonetheless, we select AM-ART to emulate memory loss phenomena due to its (i) high consistency with the neural and psychological basis in terms of both the network architecture and functional dynamics and (ii) comprehensively defined memory encoding and retrieval parameters and mechanisms.

In the perspective of AI, modelling long-term memory loss is essential towards self-regulating systems to accommodate physical memory constraints. For example, to achieve better efficiency, deep reinforcement learning agents normally perform mini-batch learning based on the experience replay strategy (Lin 1993). Other than the improvement in time-wise learning efficiency, experience replay also possesses the following perk: "the behavior distribution is averaged over many of its previous states, smoothing out learning and avoiding oscillations or divergence in the parameters" (Mnih et al. 2013). However, by performing random sampling, the conventional experience replay strategy ignores the importance or the quality of different experiences. To incorporate the quality of the experiences during sampling, various experience replay techniques, such as prioritized (Schaul et al. 2015), hindsight (Andrychowicz et al. 2017) and dual (Wei et al. 2018), have been proposed in the literature. Nonetheless, these extended strategies are built upon purely goal-orientated mechanisms, without any neurocognitive basis. Although not being the focus of this paper, it will be quite stimulating to implement autonomous agents that are able to emulate human memory recall behaviours.

AM-ART Model and Its Dynamics

The network structure of Autobiographical Memory-Adaptive Resonance Theory (AM-ART) model is shown in Figure 1. AM-ART is a three-layer neural network, wherein the event-specific knowledge of autobiographical memory is presented to the bottom layer F_1 to encode life events in the middle layer F_2 and a sequence of related events in F_2 are encoded into an episode in the top layer F_3. AM-ART is consistent with the hierarchical model established by Conway and Pleydell-Pearce (2000), which is supported by neural imaging evidence (Addis et al. 2012), in terms of both the network architecture and functional dynamics (Wang, Tan, and Miao 2016). Furthermore, we find that the circuit of AM-ART network may reside in the temporal lobe of the human brain (see Figure 1). Specifically, inputs of time and location may be from entorhinal cortex (Kraus et al. 2015), inputs of people and activity may be from fusiform gyrus (Kanwisher 2001), inputs of emotion and imagery may be from amygdala (Phelps 2004), and both the F_2 and F_3 layers may reside in hippocampus (Stark et al. 2013). Please note

that the inputs to AM-ART are considered as recognized or processed information, e.g., imagery used for memory encoding in hippocampus comes from amygdala (Phelps 2004) rather than directly from occipital lobe.

AM-ART extends the network structure of fusion ART (Tan, Carpenter, and Grossberg 2007), which is a generic self-organizing neural network comprising two layers of neural fields connected by bidirectional conditional links. However, the same bottom-up search and top-down readout operations between the layers still apply in AM-ART.

Dynamics of Fusion ART

With reference to F_1 (comprising six input channels) and F_2 (comprising one association channel) shown in Figure 1, we introduce the dynamics of fusion ART as follows.

Input vectors: Let $\mathbf{I}^k = (I_1^k, I_2^k, \ldots, I_L^k)$ denote the input vector, where I_l^k denotes input l to channel k, for $l = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, K$, where L denotes the length of $\mathbf{I}^k$ and K denotes the number of input channels.

Input channels: Let F_1^k denote an input channel that receives $\mathbf{I}^k$ and let $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_L^k)$, where $x_l^k \in [0, 1]$, denote the activation vector of F_1^k receiving $\mathbf{I}^k$. If fuzzy ART operations (see (1) and (3)) are used, $\mathbf{x}^k$ is further augmented with a complement vector $\overline{\mathbf{x}}^k$, where $\overline{x}_l^k = 1 - x_l^k$. This augmentation is named complement coding, which is applied to prevent the "code proliferation" problems (Carpenter, Grossberg, and Rosen 1991). For comprehensive discussions on complement coding and fuzzy ART operations, interested readers may refer to (Wang and Tan 2015a).

Association channel: Let $\mathbf{y} = (y_1, y_2, \ldots, y_J)$ denote the activation vector of F_2, where J denotes the number of codes in F_2. Please note that there are always $J-1$ committed (learned) codes and one uncommitted (Jth) code in F_2. If fusion ART learns from scratch, it only has one uncommitted code in F_2 (weight vector is set to 1s).

Weight vectors: Let $\mathbf{w}_j^k$ denote the weight vector of the jth code C_j in F_2 for learning the input patterns in F_1^k.

Parameters: The dynamics of fusion ART are regulated by the parameters associated with each input channel, namely choice parameters $\alpha^k > 0$, learning rate parameters $\beta^k \in [0, 1]$, contribution parameters $\gamma^k \in [0, 1]$, where $\sum \gamma^k = 1$, and vigilance parameters $\rho^k \in [0, 1]$.

Code activation: A bottom-up memory search first starts from the computation of the activation values in all codes in F_2. Specifically, given $\{\mathbf{x}^k|_{k=1}^K\}$, for each F_2 code C_j, the corresponding activation T_j is computed as follows:

$$T_j = \sum_k \gamma^k \frac{|\mathbf{x}^k \wedge \mathbf{w}_j^k|}{\alpha^k + |\mathbf{w}_j^k|}, \tag{1}$$

where the fuzzy AND operation $\wedge$ is defined by $p_i \wedge q_i \equiv min(p_i, q_i)$ and the norm $|.|$ is defined by $|\mathbf{p}| \equiv \sum_i p_i$.

Code competition: Given $\{T_j|_{j=1}^J\}$, the F_2 code with the highest activation value is named the winner, which is indexed at j^*, where $j^* = \arg\max_j T_j$.

Template matching: This template matching process checks whether resonance occurs at the winner code C_{j^*}. Specifically, the match between the input pattern and the weight vector of C_{j^*} is computed as follows:

$$M_{j^*}^k = \frac{|\mathbf{x}^k \wedge \mathbf{w}_{j^*}^k|}{|\mathbf{x}^k|}. \tag{2}$$

If C_{j^*} satisfies the vigilance criteria such that $\forall M_{j^*}^k \geq \rho^k$, a resonance occurs which leads to the subsequent learning or readout process. Otherwise, a mismatch reset occurs in which $T_{j^*} \leftarrow 0$ until a resonance occurs at another F_2 code. When an uncommitted code (definitely satisfies the criteria as weights are all 1s) is identified as the winner and recruited for learning, it becomes committed. Subsequently, a new uncommitted code will be added in F_2. As such, fusion ART self-organizes its network structure (Wang and Tan 2016).

Template learning: If learning is required, once C_{j^*} is identified, its corresponding weight vectors are updated by the following learning rule:

$$\mathbf{w}_{j^*}^{k(\text{new})} = (1 - \beta^k)\mathbf{w}_{j^*}^{k(\text{old})} + \beta^k(\mathbf{x}^k \wedge \mathbf{w}_{j^*}^{k(\text{old})}). \tag{3}$$

Knowledge readout: When this top-down knowledge readout process is invoked, C_{j^*} presents its weight vectors to the input fields, such that $\mathbf{x}^{k(\text{new})} = \mathbf{w}_{j^*}^k$.

Encoding and Retrieval of Events in AM-ART

To make the activation vectors $\mathbf{x}^k$ (5W1H of a life event) in each input channel of F_1 generic, we use normalized values to represent time (when) and location (where) and use categorical values to represent people (who), activity (what), emotion (how) and imagery (which) (all with complements).

Time vector ($\mathbf{x}^1$): It represents *when* the event happened in the form of normalized year: $x_1^1 = (I_1^1 - 1900)/200$, month: $x_2^1 = I_2^1/12$, and day: $x_3^1 = I_3^1/31$.

Location vector ($\mathbf{x}^2$): It represents *where* the event happened in the form of normalized latitude: $x_1^2 = (I_1^2 + 90)/180$ and longitude: $x_2^2 = (I_2^2 + 180)/360$ ($\mathbf{I}^2$ is determined using the Google Geocoder API).

People vector ($\mathbf{x}^3$): It is a binary-valued vector representing *who* were involved in the event. Its length corresponds to the categorization of people based on inter-personal relationship. For the dataset used in this paper, we define eight types of relationship, namely family, neighbours, spouse, friends, classmates, colleagues, acquaintances and strangers.

Activity vector ($\mathbf{x}^4$): It is a binary-valued vector representing *what* was the event. Similarly, its length corresponds to the categorization of activities. For the dataset used in this paper, we define eight types of activities, namely work, meal, sports, travel, school, shopping, religious and leisure.

Emotion vector ($\mathbf{x}^5$): It is a binary-valued vector representing *how* was the feeling during the event. Emotion is an important component of our past experience, which highly affects the encoding and retrieval of autobiographical memories (Berntsen and Rubin 2002). We categorize nine types of emotion, namely neutral, astonished, excited, happy, satisfied, tired, sad, miserable and annoyed, following the classical valence-arousal model (Russell 1980), which has been widely adopted in various computational models, e.g., (Wang and Tan 2014; Tang et al. 2017).

Algorithm 1 Event encoding and retrieval in AM-ART

encode $\mathbf{x}^k$ in F_1 w.r.t the given input pattern $\mathbf{I}^k$
activate all codes in F_2 {code activation, see (1)}
repeat
  selecting the winner code C_{j^*} {code competition}
until resonance occurs {template matching, see (2)}
if encoding is required **then**
  perform learning {template learning, see (3)}
else if retrieval is required **then**
  read out $\mathbf{w}_{j^*}^k$ in F_1 {knowledge readout}
end if

Algorithm 2 Episode encoding and retrieval in AM-ART

for all subsequent events of an episode **do**
  select the winner code C_{j^*} in F_2 w.r.t $\mathbf{x}^k$ in F_1
  $\mathbf{y}_{j^*} \leftarrow 1$, or a predefined value if using partial sequence to identify the episode
  for all previously selected codes in F_2 **do**
    $\mathbf{y}_j^{(\text{new})} \leftarrow \mathbf{y}_j^{(\text{old})}(1-\tau)$
  end for
end for
Select the winner code $j^{*\prime}$ in F_3 w.r.t $\mathbf{y}$
if encoding is required **then**
  learn the weight vector $\mathbf{w}'_{j^{*\prime}}$ in F_3:
  $\mathbf{w}_{j^{*\prime}}^{\prime(\text{new})} \leftarrow (1-\beta_2)\mathbf{w}_{j^{*\prime}}^{\prime(\text{old})} + \beta_2(\mathbf{y} \wedge \mathbf{w}_{j^{*\prime}}^{\prime(\text{old})})$
else if retrieval is required **then**
  read out $\mathbf{w}'_{j^{*\prime}}$ in F_2
end if

Imagery vector $(\mathbf{x}^6)$**:** It is a binary-valued vector representing *which* pictorial memory was associated with the event. Its value encodes the specific repository address of the stored imagery. During memory retrieval, this vector is not presented along with the others as a part of the retrieval cue. In other words, this imagery field is only involved when encoding the life events and retrieving particular pieces of memories for visual playback (Wang and Tan 2015b).

The F_2 layer of AM-ART encodes events. The process of event encoding and retrieval is shown in Algorithm 1.

Encoding and Retrieval of Episodes in AM-ART

Assume the related events of one episode happened at $t_0, t_1, \ldots, t_n$ and let y_{t_i} denote the activation value of the event happened at t_i. To encode the sequence of the events, we need to always hold the inequality that $y_{t_n} > y_{t_{n-1}} > \cdots > y_{t_0}$. Therefore, we use a succession parameter $\tau \in (0,1)$ to regulate the activation sequence, such that $y_j^{(\text{new})} = y_j^{(\text{old})}(1-\tau)$ at each new time step. The F_3 layer of AM-ART encodes episodes to associate the related events encoded in F_2. The process of episode encoding and retrieval is shown in Algorithm 2.

Memory Loss during Formation

During the memory formation process, memory loss occurs in the form of encoding failure, which is caused by the deactivation of certain brain region(s) due to a demanding cognitive task (Daselaar et al. 2009). Therefore, in AM-ART, we introduce the overload parameter $\lambda \in [0,1]$ to regulate the likelihood of one being affected by cognitive overload during memory formation. Specifically, λ influences the vigilance parameters ρ^k (see *template matching*) and contribution parameters γ^k (see *code activation*) as follows:

$$\rho^k = \begin{cases} 1-\lambda(1-\textit{rand}()^k), & \text{if } \textit{rand}()^k > \lambda, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where $\textit{rand}() \in [0,1]$ generates a random number and

$$\gamma^k = \frac{\rho^k}{\sum_k \rho^k}. \quad (5)$$

Due to the lack of quantitative studies in the related neurobiology and neurocognitive literature, there does not exist a good reference on how to determine the cognitive load during memory formation based on both one's state of mind and external stimuli. Instead, we have to employ a random generator $\textit{rand}()^k$ to emulate the cognitive capability on the kth input channel in F_1 during the formation of each life event. Thus, equation (4) describes that with probability λ (if $\textit{rand}()^k \leq \lambda$), the kth input channel is overlooked ($\rho^k = \gamma^k = 0$) during memory formation due to cognitive overload in the respective brain region. Otherwise, the vigilance equals to the level of attention, which is estimated as $1-\lambda(1-\textit{rand}()^k)$, i.e., a lower λ value and a higher $\textit{rand}()^k$ value lead to the formation of more distinguishable memory.

The process of memory loss during formation is shown in Algorithm 3. Generally speaking, people in different life stages, denoted as t_i, differ in λ_{t_i}. In our emulations, we learn the values of λ_{t_i} using published memory survey data.

Algorithm 3 Memory loss process during formation

upon receiving an input pattern $\mathbf{I}^k$ for memory formation, encode $\mathbf{x}^k$ in F_1, furthermore, update ρ^k and γ^k {overload effect, see (4) and (5), respectively}
identify the winner code C_{j^*} where resonance occurs
perform encoding for memory formation
if memory encoded is the first in a new time period, $\forall j \neq j^*$ in F_2, decrease v_j {decay effect, see (6)}

Memory Loss during Storage

During long-term storage, memory decays along time due to inactivation. Although this decay is monotonic, its rate declines rapidly at first and then much more slowly, which well fits an exponential curve (Rubin 1982). Therefore, in AM-ART, we introduce the decay parameter $\phi \in [0,1]$ to regulate the rate of long-term memory fading. Moreover, we introduce the vividness parameter $v_j \in [0,1]$, which associates with each event in F_2 to denote the vividness of the jth event. Upon encoding (see (3)) at t_a, event j has the highest level of vividness, i.e., $v_j = 1$. Specifically, as time elapses, the vividness of an encoded event decays (see Step 4 of Algorithm 3) in the following manner:

$$v_j^{(new)} = \max(0, v_j^{(old)} - \exp(\phi - (t_i - t_a))), \text{ if } i > a, \quad (6)$$

where $\exp(\phi-(t_i-t_a))$ denotes the decay rate and (t_i-t_a) denotes the amount of elapsed time. Because $\phi \leq 1$ and $t_i-t_a \geq 1$, the decay rate is nicely bounded within the $[0,1]$ interval. When $v_j \leq 0$, the jth event is no longer retrievable.

On the other hand, for healthy persons, their memory gets refreshed through reactivation (Gisquet-Verrier and Riccio 2012), wherein a similar pattern of the associated features is recalled (Chalfonte and Johnson 1996). Therefore, during memory retrieval, the vividness of a winner event j increases proportionally to its activation value (see (1)) due to reactivation (see Step 4 of Algorithm 4) in the following manner:

$$v_j^{(new)} = \min(1, v_j^{(old)} + \exp(\phi-(t_i-t_a))T_j), \text{ if } v_j^{(old)} > 0. \quad (7)$$

The decay rate can be rewritten as $\exp(\phi)\exp(-(t_i - t_a))$, which means a higher ϕ value and longer elapsed time lead to greater memory decay or reactivation. In our emulations, we learn the values of ϕ_{t_i} associated with different life stages using published memory survey data.

Memory Loss during Retrieval

Memory loss during retrieval manifests as retrieval-induced forgetting (RIF), which refers to the phenomenon of certain information becomes less recallable due to memory interdependency (Storm and Levy 2012). RIF has been identified as goal-directed and may not necessarily within conscious control (Barnier, Hung, and Conway 2004). Among the various possible intricate accounts of memory retrieval inhibition, we adopt the two prominent ones that have been most widely and frequently supported by empirical studies (Storm and Levy 2012), namely *cue independence*, which means RIF takes place regardless of the choice of retrieval cues, and *competition dependence*, which means RIF is affected by the similarity between the to-be-retrieved piece of memory and its competitors. Furthermore, although it might be well-known that the elderly tend to recall more positive memories than negative ones, emotional inhibition has been identified in young adults as well (Barnier, Hung, and Conway 2004). Therefore, in AM-ART, we introduce the inhibition parameter $\mu \in [0,1]$ to regulate the likelihood of retrieval failure. Specifically, when the jth event in F_2 is identified as the winner, before checking whether resonance occurs, its activation value may be reset due to inhibition, which is regulated in the following manner:

$$T_j = 0, \text{ if } \textit{rand}() < \mu(1-(T_j - T_l)T_j)\zeta_{t_i}^s. \quad (8)$$

where l denotes the index of the event that has the second highest activation value (see (1)) and $\zeta_{t_i}^s$ denotes one's emotional coefficient parameter in life stage t_i associated with affective state s of the winner event j. It is obvious in (8) that with a larger activation value T_i of the winner event and a larger difference between the winner and the runner-up $(T_j - T_l)$, the chance of the winner gets inhibited from retrieval is smaller. Moreover, the value of $\zeta_{t_i}^s$ is bounded between that of the most negative state $\zeta_{t_i}^-$ (low valence and low arousal, see the 2-D circumplex model of affect (Russell 1980)) and that of the most positive state $\zeta_{t_i}^{++}$ (high valence and high arousal), which can be computed as follows:

$$\zeta_{t_i}^s = \zeta_{t_i}^- + \frac{1+\cos(\theta^s - 45^\circ)}{2}(\zeta_{t_i}^{++} - \zeta_{t_i}^-), \quad (9)$$

Algorithm 4 Memory loss process during retrieval

1: upon receiving a cue $\mathbf{U}^k$ for memory retrieval, encode $\mathbf{x}^k$ in F_1, furthermore, reset $\rho^k \leftarrow \rho_0^k$ and $\gamma^k \leftarrow \gamma_0^k$
2: **repeat**
3: selecting the winner code C_{j^*} for inhibition check
4: increase v_{j^*} {reactivation effect, see (7)}
5: **if** inhibition occurs **then**
6: $T_{j^*} \leftarrow 0$ {inhibition effect, see (8)}
7: **else**
8: further check if resonance occurs at C_{j^*}
9: **end if**
10: **until** resonance occurs
11: perform readout for memory retrieval

where $\theta^s \in [0^\circ, 360^\circ]$ denotes the angle of affective state s in the 2-D circumplex. Moreover, $\zeta_{t_i}^{neutral} = \frac{1}{2}(\zeta_{t_i}^{++} + \zeta_{t_i}^-)$. In our emulations, we learn the values of μ_{t_i}, $\zeta_{t_i}^-$ and $\zeta_{t_i}^{++}$ using published memory survey data.

The memory loss process during retrieval is shown in Algorithm 4, wherein the initial parameter values of AM-ART are denoted with 0 in the subscript. Unlike memory loss during storage that an event can no longer be retrieved once its vividness decreases to zero (see (6)), RIF only causes the memory temporarily inaccessible to conscious recall (Barnier, Hung, and Conway 2004).

Using AM-ART to Model Memory Loss

To validate our approach of using AM-ART to model memory loss, we collect a memory dataset from public domains and use it to conduct all the experiments in this paper. Please note that the collection of a relatively large real-world autobiographical memory dataset is definitely necessary because the natural relationships among the event-specific knowledge (Conway and Pleydell-Pearce 2000) reflect actual scenarios and remain relatively consistent throughout one's life, which a randomly generated memory set cannot offer.

Our collected dataset comprises 1,019 snapshots of life events (5W1H) in 131 episodes of Mr. Obama, the 44th President of USA. Other persons' memory sets can also be used for the experiments conducted in this paper. We simply choose Mr. Obama because his life events are largely available online with rich context (for tagging 5W1H) and no privacy issue is involved. Specifically, we directly extract the images and their corresponding context from the online web pages (Zimbio.com and Google Images) except emotion (manually derived, as emotion recognition based on image and its context is not the focus of this paper). However, because this dataset does not evenly span across one's entire life (less memory collection in childhood and young adulthood), we alter the event dates and (roughly) equally distribute the events across all life stages based on the intuitive assumption that the number of events experienced during the same length of long time periods should also be equal. The number of life stages is set to eight, which follows the categorization criteria used by Berntsen and Rubin (2002) that from 0s to 70s, each has ten years' time span, i.e., $t_i \in \{0, 1, \ldots, 7\}$ (see (6)). Moreover, we make sure the

Table 1: List of estimated emotional coefficient values.

Age	0s	10s	20s	30s	40s	50s	60s	70s
$\zeta_{t_i}^{-}$	0.431	0.547	0.625	0.644	0.615	0.632	0.449	0.352
$\zeta_{t_i}^{++}$	0.877	0.893	0.882	0.900	0.842	0.805	0.716	0.671

Table 2: List of initial parameter values used in experiments.

Parameter	**Value**	**Description/Remark**
Choice (α_0^k)	0.001	Mainly used to avoid having NaN in (1)
Learning rate (β_0^k)	0.5	Not in use during memory retrieval
Contribution (γ_0^k)	0.167	Equally assigned, such that $\sum \gamma_0^k = 1$
Vigilance (ρ_0^k)	0.9	During memory formation, determined by (4)
Succession rate (τ)	0.05	Used for encoding event sequence (see Algo. 2)

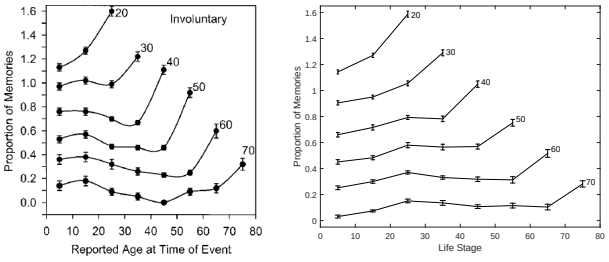
ratio among pleasant ($\cos(\theta^s - 45°) > 0$, see (9)), neutral and unpleasant ($\cos(\theta^s - 45°) < 0$) memories in each life stage conforms to the distribution reported in Figure 12 of (Berntsen and Rubin 2002), in which a significantly higher ratio of pleasant memories is reported in 20s. Please note that when an episode is selected for date alternation, the dates of all its events are changed to the corresponding life stage, following the original event sequence. As such, although certain episode sequence may become unnatural in real world, this necessary event date alteration procedure does not affect the utility of our proposed model.

Furthermore, before we conduct the memory loss emulations, we predetermine the emotional coefficient parameter value ranges (see (9)) based on the number of emotional memory recalls reported in Table 1 of (Rubin and Berntsen 2003), which extends their prior study (Berntsen and Rubin 2002) (more human subjects: 1,307 VS 1,241). Specifically, we compute $\zeta_{t_i}^{++}$ using the ratio of the total number of "pride" and "love" memory recalls over the total number of their attempts. Similarly, we compute $\zeta_{t_i}^{-}$ based on "fear", "jealousy" and "anger". Please note that "important" memory recalls listed in the same table is not used as they do not tie to any particular emotion. The predetermined parameter values are reported in Table 1. Because the memory recalls in 0s and 10s are missing from (Rubin and Berntsen 2003), we estimate those values by polynomial extrapolating the same emotional coefficient values in other age groups. We set the polynomial degree to 2 because it is low enough to avoid overfitting and high enough to well keep the extrapolated values within a certain range. For example, in Table 1, ζ_0^{++} will be greater than 1 if the polynomial degree is set to 1 (i.e., linear extrapolation).

The initial parameter values of AM-ART are listed in Table 2. Most such parameters take on a standard set of parameter values and all do not require tuning during runtime.

Emulating Memory Loss Across One's Life Span

In the study conducted by Berntsen and Rubin (2002), they interviewed 1,241 subjects aged 20 and above to learn their memory recalls across their life span in various manners. Among the various assessments they reported, involuntary autobiographical memory recalls may best represent the distribution of the well-preserved memories across one's life



(a) (b)

Figure 2: Memory recall distributions of different age groups across life span: (a) Figure 6 of [Berntsen and Rubin, 2002]. (b) Results of AM-ART emulations. To make all plots visible, an offset of 0.2 is applied to each adjacent age group.

span (see Figure 2(a)). Therefore, we use this set of reported proportion of memories recalled in different life stages to investigate the following research question:

How accurately can our proposed computational memory loss procedures replicate the memory recall behavioural patterns observed in real world?

To answer the above question by applying AM-ART memory loss procedures on the survey data visualized in Figure 2(a), we need to assume that an individual's memory loss parameter values do not vary within the same life stage, i.e., λ_{t_i}, ϕ_{t_i} and μ_{t_i} associated with each individual, where $t_i \in \{0, 1, \ldots, 7\}$, remain invariant. Furthermore, we use Genetic Algorithm (GA) (Goldberg 1989) to emulate the individual subjects (assume they all went through the same life events at each life stage) and minimise the difference (root mean square error, RMSE) between the emulated memory recall performance and the published survey data. Specifically, for each age group, the chromosome length is set to $3\times(t_i+1)$ and each gene represents one of λ_{t_i}, ϕ_{t_i} and μ_{t_i} in real number. The various GA strategies employed are tournament selection of parents (size=2 and probability=0.75), uniform crossover (rate=1), bounded mutation (to ensure all gene values are kept within $[0, 1]$, rate=0.75), and elitism replacement (ratio=0.1). For each age group, the population size is set to 200 and GA terminates after 20 iterations. In addition, we maintain a pool of identical best-performers across GA iterations in parallel. The pool size is set to 200, which is close to the averaged number of subjects in each age group (206.83) interviewed by Berntsen and Rubin (2002).

The emulations are conducted as follows. For each age group, each individual and each life stage, memory formation (encoding) first takes place. Upon proceeding to the subsequent life stage, memory decay takes place. Moreover, all the prior memories are used once again as retrieval cues to emulate memory rehearsal (reactivation). In the end, one's retrieval performance at the last life stage is recorded as the final emulation result. The performance of the 200 individuals kept in the pool is averaged and visualized in Figure 2(b).

The curves shown in Figure 2(b) appear to be more stable, but are highly consistent with those in Figure 2(a) that the averaged correlation of all the age groups across

Table 3: Prediction errors of the estimated parameter values.

Prediction of	30s	40s	50s	60s	70s
Linear	**0.0119**	**0.0141**	0.0312	0.0495	0.0732
Polynomial	0.0177	0.0169	**0.0294**	**0.0363**	**0.0516**
Random	0.0231	0.0260	0.0358	0.0686	0.0822

Note: Polynomial degree=2; Random means a parameter value is randomly generated from the [min, max] range of the corresponding values in previous life stages.

each life stage between these two subfigures is computed as 0.793 ± 0.166. Moreover, the phenomena observed in Figure 2(b) are highly consistent with the widely reported literature that "older adults demonstrate a three-component pattern in the distribution of memories across the life span: few memories from childhood (childhood amnesia), a bump in young adulthood followed by a decrease in midlife (a reminiscence bump), and increase in later years (a recency effect)" (Fromholt et al. 2003). Although both subfigures show the *reminiscence bump* widely observed "between the ages of 10 and 30" (Demiray, Gulgoz, and Bluck 2009), the bump in Figure 2(a) is in 10s while that in Figure 2(b) is in 20s. This difference may be explained by the fact that Figure 12 of (Berntsen and Rubin 2002), which comprehensively visualizes the distribution of emotional memories across all eight life stages, actually reports the re-analysed distribution of another population (Rubin and Schulkind 1997) (see Figure 12's caption of (Berntsen and Rubin 2002)).

Moreover, we find that the results shown in Figure 2(b) are highly consistent with comparable memory assessments reported in another well-known study (Rubin and Schulkind 1997), wherein the memory recall ratios of 20s, 30s and 70s within past two decades (see Table 2 of (Rubin and Schulkind 1997)) are computed as 0.830, 0.776 and 0.476, respectively. These ratios are remarkably similar to the corresponding AM-ART emulation results of 0.857, 0.745 and 0.386, respectively.

Predicting Memory Loss in Subsequent Life Stage

We further test whether the learnt parameter values can be used to predict memory loss in one's subsequent life stage. Specifically, we extrapolate the learnt parameter values of age group t_i to predict their memory performance in t_{i+1}. The prediction results in terms of RMSE based on the 200 best-performing individuals are reported in Table 3. As shown, when predicting one's memory performance in the latter life stages ($t_i \geq 5$), polynomial extrapolation performs much better than linear extrapolation. This finding is consistent with the widely reported literature that elderly's memory performance declines rapidly as they age (Small et al. 1999; Wang et al. 2014; Wang and Tan 2017).

Applicability of Modelling Memory Loss in Agents

Being able to model long-term memory loss like human does may shed light upon the design of memory consolidation and utilization strategies in autonomous agents. For example, our memory loss model can be straightforwardly employed by a deep reinforcement learning agent with limited memory capacity in a complex game environment to effectively select diverse experiences to be preserved for batch learning. Such linkage between our human memory loss model and the agent's memory discard strategy will be quite stimulating that an agent is enabled to emulate human's memory recall behaviours, e.g., better preservation of recent (adaptivity in the case of agent), happy (higher rewards), and young-adulthood (most frequently referenced) memories.

Conclusion

In this paper, we introduce the dynamics of a neurocomputational autobiographical memory model on how to replicate real-world memory loss phenomena based on well established neurocognitive theories. The emulation results show high correlation with human memory recall performance. Although our approach may only replicate one of the many possible mechanisms used by human brain, it can be considered as a piece of ground-breaking work in this research direction. Going forward, we will implement the stimulating memory discard strategy in autonomous agents to investigate the implications of their human-like behaviours.

Acknowledgments

This research is supported, in part, by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative and the Singapore Ministry of Health under its National Innovation Challenge on Active and Confident Ageing (NIC Project No. MOH/NIC/COG04/2017 and MOH/NIC/HAIG03/2017).

References

Addis, D. R.; Knapp, K.; Roberts, R. P.; and Schacter, D. L. 2012. Routes to the past: Neural substrates of direct and generative autobiographical memory retrieval. *Neuroimage* 59(3):2908–2922.

Anderson, J. R.; Bothell, D.; Byrne, M. D.; Douglass, S.; Lebiere, C.; and Qin, Y. 2004. An integrated theory of the mind. *Psychological Review* 111:1036–1060.

Andrychowicz, M.; Crow, D.; Ray, A.; Schneider, J.; Fong, R.; Welinder, P.; McGrew, B.; Tobin, J.; Abbeel, P.; and Zaremba, W. 2017. Hindsight experience replay. In *NIPS*, 5055–5065.

Barnier, A. J.; Hung, L.; and Conway, M. A. 2004. Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion* 18(4):457–477.

Berntsen, D., and Rubin, D. C. 2002. Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic, and involuntary memories. *Psychology and Aging* 17(4):636–652.

Carpenter, G. A.; Grossberg, S.; and Rosen, D. B. 1991. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks* 4(6):759–771.

Chalfonte, B. L., and Johnson, M. K. 1996. Feature memory and binding in young and older adults. *Memory and Cognition* 24(4):403–416.

Conway, M. A., and Pleydell-Pearce, C. W. 2000. The construction of autobiographical memories in the self-memory system. *Psychological Review* 107(2):261–288.

Daselaar, S. M.; Prince, S. E.; Dennis, N. A.; Hayes, S. M.; Kim, H.; and Cabeza, R. 2009. Posterior midline and ventral parietal activity is associated with retrieval success and encoding failure. *Frontiers in Human Neuroscience* 3:1–10.

Demiray, B.; Gulgoz, S.; and Bluck, S. 2009. Examining the life story account of the reminiscence bump: Why we remember more from young adulthood. *Memory* 17(7):708–723.

Derbinsky, N., and Laird, J. E. 2013. Effective and efficient forgetting of learned knowledge in Soar's working and procedural memories. *Cognitive Systems Research* 24:104–113.

Fromholt, P.; Mortensen, D. B.; Torpdahl, P.; Bender, L.; Larsen, P.; and Rubin, D. C. 2003. Life-narrative and word-cued autobiographical memories in centenarians: Comparisons with 80-year-old control, depressed, and dementia groups. *Memory* 11(1):81–88.

Gauthier, I.; McGugin, R. W.; Richler, J. J.; Herzmann, G.; Speegle, M.; and VanGulick, A. E. 2014. Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision* 14(8):article 7.

Gisquet-Verrier, P., and Riccio, D. C. 2012. Memory reactivation effects independent of reconsolidation. *Learning and Memory* 19:401–409.

Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

Jahn, H. 2013. Memory loss in Alzheimer's disease. *Dialogues in Clinical Neuroscience* 15:445–454.

Kanwisher, N. 2001. Neural events and perceptual awareness. *Cognition* 79:89–113.

Kraus, B. J.; Brandon, M. P.; Robinson II, R. J.; Connerney, M. A.; Hasselmo, M. E.; and Eichenbaum, H. 2015. During running in place, grid cells integrate elapsed time and distance run. *Neuron* 88(3):578–589.

Laird, J. E. 2012. *The Soar Cognitive Architecture*. MIT Press.

Langley, P. 2006. Cognitive architectures and general intelligent systems. *AI Magazine* 27:33–44.

Lin, L.-J. 1993. *Reinforcement learning for robots using neural networks*. Ph.D. Dissertation, Carnegie Mellon University.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing Atari with deep reinforcement learning. arXiv:1312.5602.

Phelps, E. A. 2004. Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology* 14:198–202.

Rubin, D. C., and Berntsen, D. 2003. Life scripts help to maintain autobiographical memories of highly positive, but not highly negative, events. *Memory and Cognition* 31(1):1–14.

Rubin, D. C., and Schulkind, M. D. 1997. Distribution of important and word-cued autobiographical memories in 20-, 35-, and 70-year-old adults. *Psychology and Aging* 12(3):524–535.

Rubin, D. C. 1982. On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior* 21(1):21–38.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6):1161–1178.

Schaul, T.; Quan, J.; Antonoglou, I.; and Silver, D. 2015. Prioritized experience replay. arXiv:1511.05952.

Small, S. A.; Stern, Y.; Tang, M.; and Mayeux, R. 1999. Selective decline in memory function among healthy elderly. *Neurology* 52(7):1392–1396.

Stark, S. M.; Yassa, M. A.; Lacy, J. W.; and Stark, C. E. L. 2013. A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia* 51(12):2442–2449.

Storm, B. C., and Levy, B. J. 2012. A progress report on the inhibitory account of retrieval-induced forgetting. *Memory and Cognition* 40(6):827–843.

Tan, A.-H.; Carpenter, G. A.; and Grossberg, S. 2007. Intelligence through interaction: Towards a unified theory for learning. In *ISNN*, 1094–1103.

Tang, C.; Wang, D.; Tan, A.-H.; and Miao, C. 2017. EEG-based emotion recognition via fast and robust feature smoothing. In *BI*, 83–92.

Wang, D., and Tan, A.-H. 2014. Mobile humanoid agent with mood awareness for elderly care. In *IJCNN*, 1549–1556.

Wang, D., and Tan, A.-H. 2015a. Creating autonomous adaptive agent in a real-time first-person shooter computer game. *IEEE Transactions on Computational Intelligence and AI in Games* 7(2):123–138.

Wang, D., and Tan, A.-H. 2015b. MyLife: An online personal memory album. In *WI-IAT*, 243–244.

Wang, D., and Tan, A.-H. 2016. Self-regulated incremental clustering with focused preferences. In *IJCNN*, 1297–1304.

Wang, D., and Tan, A.-H. 2017. eHealthPortal: A social support hub for the active living of the elderly. In *ICCSE*, 19–25.

Wang, D.; Subagdja, B.; Kang, Y.; Tan, A.-H.; and Zhang, D. 2014. Towards intelligent caring agents for aging-in-place: Issues and challenges. In *CIHLI*, 1–8.

Wang, P.; Gauthier, I.; and Cottrell, G. 2016. Are face and object recognition independent? A neurocomputational modeling exploration. *Journal of Cognitive Neuroscience* 28(4):558–574.

Wang, D.; Tan, A.-H.; and Miao, C. 2016. Modelling autobiographical memory in human-like autonomous agents. In *AAMAS*, 845–853.

Wei, Z.; Wang, D.; Zhang, M.; Tan, A.-H.; Miao, C.; and Zhou, Y. 2018. Autonomous agents in Snake game via deep reinforcement learning. In *ICA*, 20–25.