Technical Report: Modeling the Occurrence of Intrusive Memories in Post-Traumatic Stress Disorder

Briana Smith (brianam2@uw.edu)
Department of Neuroscience, Campus Box 351525
University of Washington, Seattle, WA 98195

Andrea Stocco (stocco@uw.edu)
Department of Psychology, Campus Box 351525
University of Washington, Seattle, WA 98195

Abstract

This document presents a technical description of a model of intrusive memories in Post-Traumatic Stress Disorder.

Keywords: Post-Traumatic Stress Disorder; Memory; ACT-R; Rational Analysis; Computational Psychiatry

Introduction

One of the most important phenotypes of Post-Traumatic Stress Disorder (PTSD) is the existence of intrusive memories, i.e. traumatic memories whose unwanted recollection disrupts normal function and causes additional stress. Understanding the nature of intrusive memories is important for predicting recovery trajectories, which remain baffling in PTSD research.

The goal of this project is to use the tools of computational cognitive neuroscience to model and predict the surgeance of intrusive memories over time. The central idea is that the frequency of intrusive memories can be predicted given model parameters that can be estimated from individual and environmental factors.

The central idea of this project is that traumatic memories can be understood within the context of a general theory of declarative (and, specifically, episodic) memory. Theories of declarative memory are well specified computationally and have an clear interpretation in terms of biology, and thus provide an excellent reference frame to understand and model traumatic memories.

A paradoxical consequence of this assumptions is that traumatic memories are not "exceptional" *per se*; rather, their exceptionality arises as a function of specific parameters that make them behave in an an exceptional way. This is paradoxical because, while models of declarative memories stress the rationality and adaptiveness of the rules of remembering and forgetting (Anderson, 1990), intrusive memories are clearly disruptive and maladaptive. As the proposed model can show, this paradox is only apparent: intrusive memories do use the same rules that would normally be adaptive, but their exceptional qualities lead normally adaptive rules of forgetting to exceptional and chaotic behaviors. To steal the title of Redish's landmark paper on addiction (?, ?), intrusive memories are a "computational process gone awry".

The ultimate goal of this project is to use this model to precisely predict the effects of trauma over time for any individual, and possibly to reverse-engineer the best treatment.

The Model

The starting point of this model is Anderson's analysis of human episodic memory in terms of "Rational Analysis" (Anderson, 1990), or, as it would be called currently, Bayesian terms. In rational or Bayesian frameworks, a memory i's probability of being recalled in the face of a context $Q = q_1, q_2, \ldots q_n$ reflects the memorys retrieval *need*, and is a Bayesian function of both the past history of a memory and the degree to which the each individual cue $q_1, q_2 \ldots q_n$ is predictive of i.

In Anderson's classic formulation (Anderson, 1990), the need odds $N(i)$ of a memory i in a context Q are simply the posterior odds $P(i|Q)/P(\neg i|Q)$, and can be expressed as:

$$\frac{P(i \mid Q)}{P(\neg i \mid Q)} = \frac{P(i)}{P(\neg i)} \times \prod_q \frac{P(q \mid i)}{P(q)} \tag{1}$$

This equation be further analyzed to derive a mechanistic model of memory. In Anderson's further elaborations of the theory (Anderson, 2009), a memory i's need probability is expressed in terms of activation A_i, a scalar meta-quantity that represents the log of the odds:

$$\begin{aligned} A_i &= \log \frac{P(i \mid Q)}{P(\neg i \mid Q)} \\ &= \log \left(\frac{P(i)}{P(\neg i)} \times \prod_q \frac{P(q \mid i)}{P(q)} \right) \\ &= \log \left(\frac{P(i)}{P(\neg i)} \right) + \log \left(\prod_q \frac{P(q \mid i)}{P(q)} \right) \\ &= \log \left(\frac{P(i)}{P(\neg i)} \right) + \sum_q \log \left(\frac{P(q \mid i)}{P(q)} \right) \end{aligned} \tag{2}$$

In ACT-R, it is customary to give different names to the two quantities that make up the right-hand side of Eq. 2, referring to first one as the *base-level activation* of i or B_i, and to the second one as the *spreading activation* of i or S_i. Thus:

$$B_i = \log\left(\frac{P(i)}{P(\neg i)}\right) \tag{3}$$

$$S_i = \sum_q \log\left(\frac{P(q \mid i)}{P(q)}\right) \tag{4}$$

Combining Eq. 2 with Eq. 3 and 4, we have

$$A_i = B_i + S_i \tag{5}$$

Algorithmic Implementations of B_i and S_i

In algorithmic implementations of the theory, the quantities B_i and S_i are approximated in ways that predict future use (that is, the need probability $N(i)$) based on previous history (B_i) or structural similarity (S_i).

Specifically, B_i is approximated as the sum of decaying traces of i's retrievals, thus combining the number of times i has been retrieved with (frequency) with the passing amount of time (recency):

$$B_i = \log\left(\sum_{0\leq j\leq R} (t-t_j)^{-d}\right) \tag{6}$$

In Eq. 6, R represents the number of retrievals and d is a *decay* rate parameter.

Spreading activation S_i, on the other hand, is interpreted with reference to a classic representation format for memories, that is, semantic networks. In this representation, each memory represents node in a multi-dimensional space. Elements that are included in that memory are represented as directional link between the node and each of the elements. Spreading activation S_i is implemented as the amount of activation that flows from the memory nodes that are part of the context $Q = q_1, q_2 \ldots q_{N_Q}$. Thus, each element q_j represents a specific active node in the current context. If there is a direct link from q to memory i, than i receives an activation boost that is proportional the product between the strength of the link connecting q to i (indicated as sq,i) and an *attentional weight* w. The weight is usually simplified as a single scalar quantity, W, divided over the number of active elements in the context, N_Q, so that $w = W/N_Q$. The total amount of spreading activation S_i is the sum of all of partial effects of each element q:

$$\begin{aligned} S_i &= \sum_q \frac{W}{N_Q} \times s_{j,i} \\ &= \frac{W}{N_Q} \sum_q s_{j,i} \end{aligned} \tag{7}$$

Which leads to algorithmic implementation of Eq. 2 as:

$$A_i = \log\left(\sum_j (t-t_j)^{-d}\right) + \frac{W}{N_Q}\sum_q s_{q,i} \tag{8}$$

Extending the Classic Approach

It has been noted several times that the classic approach does not consider emotions, which are know to have an effect on declarative memory retrieval. A possible reason for this is that emotion is not easy or immediately incorporated in the rational theory of retrieval.

The goal of this model is to integrate emotion into the rational or Bayesian theory of memory. The simplest way to do is to consider emotions from an evolutionary perspective – what are they for? A common theme in emotion neuroscience is that emotions are needed for survival. Events that are associated with different emotions, or to the same emotion but to a different degree, also differ in *survival value*.

Although the idea that events have different survival value is obvious and mundane, it does not have any effect on memory according to the classical interpretation above. In fact, in the classic approach all memories are exactly equal, independently of their contents or the specific situations of their encoding.

A simple way to overcome this limitation is by explicitly supplying the memory's survival value as a quantity V_i. In this formulation, V_i is quantity between 0 and ∞, and a memory's need odds $N(i)$ can be generalized as the posterior odds scaled by the survival value associated with retrieving the memory itself.

The additional "survival odds" can be calculated as follows. If we imagine that all memories have exactly the same posterior odds, then the only metric of interest is their relative survival value. This can be calculated as $V(i)/V(\neg i)$, or the ratio between i's survival value and the survival value if i not chosen. In turn, $V(\neg i)$ can be thought of as the mean survival value of all other memories that are not i. As the number of memories grows, $V(\neg i)$ can be approximated as the mean survival value $\overline{V}$.

Specifically, assuming that all memories have an associated survival value expressed in any consistent metric, the need probability for a memory i can be expressed as follows:

$$\begin{aligned} N(i) &= \frac{P(i \mid Q)}{P(\neg i \mid Q)} \times \frac{V(i)}{V(\neg i)} \\ &= \frac{P(i)}{P(\neg i)} \times \prod_q \frac{P(q \mid i)}{P(q)} \times \frac{V(i)}{V(\neg i)} \\ &= \frac{P(i)}{P(\neg i)} \times \prod_q \frac{P(q \mid i)}{P(q)} \times \frac{V(i)}{\overline{V}} \end{aligned} \tag{9}$$

if we express the new version of $N(i)$ in terms of log odds, as in Eq. 2, we have:

$$\begin{aligned} A_i &= B_i + S_i + \log\left(\frac{V(i)}{\overline{V}}\right) \\ A_i &= B_i + S_i + \log V(i) - \log \overline{V} \end{aligned} \tag{10}$$

Since the term $\log \overline{V}$ is a constant bias and does not depend on i, it will be ignored from now on.

According

Relationship to Other Approaches

The idea of that emotions modulate declarative memory is not new. Perhaps the earliest attempt was made by Fum and Stocco (Fum & Stocco, 2004) in a model designed to replicate Antonio Damasio's findings with the Iowa Gambling Task (Bechara, Damasio, Tranel, & Damasio, 1997). In their model, an emotional term (also named V) is added to the computation of the spreading activation term S_i. Specifically, Eq. 7 was changed so that $S_i = \frac{W}{N_Q} \sum_q (s_{j,i} + V(i))$.

More recently, Juvina and colleagues (Juvina, Larue, & Hough, 2018) have proposed a similar model of emotion, in which the baseline activation is augmented with two terms, V and A, that correspond to a memory's emotional *valence* and *arousal*, respectively, so that a memory's full activation was defined as $B_i + S_i + V_i + A_i$.

The solution proposed here is similar to these attempts, in that it also consists in adding an additional term to the activation equation. It differs from previous attempts because it derives its expression from the original Bayesian framework of Anderson (Anderson, 1990), rather than adding terms on the bases of either neurobiology or the need to fit the data.

That being said, the current model *does* fit some basic aspects of emotion neurobiology, as the next section will show.

Biological Interpretation of the Model

The proposed model can be given an interpretation in terms of neurobiology–specifically, the neurobiology of the hippocampal circuit.

The *hippocampus* (part of the medial temporal lobe) is currently understood as the basic circuit that encodes declarative memories. Researchers still debate whether the hippocampus is also responsible for the long-term storage of memories, with some proponents suggesting that it might be, at least for episodic memories. Either way, it remains the one circuit responsible for originally creating memories.

The hippocampus receive topologically organized projections from all over the cortex, and, in turn, send projections back. Internally, hippocampal neurons form an interconnected network. It has been suggested multiple times that this particular pattern of connectivity can be interpreted as follows: the hippocampus works as an autoassociator or an autoencoder, storing patterns of cortical activity that can then be re-created based on partial inputs.

The *amygdala* is a small nucleus located in front of the hippocampus. It plays a fundamental role in processing emotions, particularly fear and stress. Like the hippocampus, it receives widespread cortical projections. The amygdala projects directly to the hippocampus, but does not receive projections from it.

The individual components of this circuit can be mapped onto the model:

- The decaying term B_i reflects intrinsic activity of the hippocampal autoassociator. This is an old idea, tracing back to Alvarez and Squire's model. It can also be compared to the multiple-trace model.
- the spreading activation term $\sum_q W_{i,q} A_q$ reflects the activity of cortical inputs through the dentate gyrus, providing the contextual inputs to memory retrieval.
- The terms $V(i)$ reflects the contribution of the amygdala.

The new term $V(i)$ can be also be interpreted in reference to the literature on emotional processing. Although authors disagree on hw to categorize emotions, most authors agree that emotions can be placed at least across two dimensions, it valence and *arousal*. The term $V(i)$ can be thought of as the norm of the vector in this bidimensional space, capturing an emotion's survival value into a single metric.

Model Implementation

An algorithmic version of the model was implemented in the ACT-R cognitive architecture [1]. ACT-R is a natural choice, since it's declarative memory system reflects the Rational Analysis framework that is the basis of Eq. 8. Thus, only minor modifications were needed to implement the new activation equation (Eq. 10).

Although ACT-R is a complex architecture with detailed assumptions about cognitive, perceptual, and motor functions, only the declarative memory portion was used in these simulations.

Model Behavior

The model simulates behavior over an extended period of time, somewhat stretching the boundaries of ACT-R's typical timescale. The model follows a simple perceive-retrieve-respond loop (as illustrated in Algorithm 1). At predefined and fixed intervals of time (ΔT), the model is presented with a new situation, which is represented as new chunk C. The model responds to the new situation by setting a goal to resolve it. When the goal is set, the model retrieves the most active memory i. Once the memory is retrieved, the goal is resolved.

Note that, since activation A_i would reflect the need probability $N(i)$, under normal circumstances, the retrieved memory i would the most needed and, therefore, the most relevant to the current situation.

At a predefined time T_{PTE}, a *potentially traumatic event* (PTE) is generated and presented to the model. This event is marked as (potentially) having a greater-than-normal emotional impact.

Representation of Memories

In ACT-R, memories are represented as "chunks", which are essentially vector-like structures containing a predefined number of *slots*, each of which contains one *attribute* . In the model, new memories are arbitrary and randomly generated. The structures of these memories is controlled by two parameters, N_S and N_A, which determine the number of slots in a given memory (N_S, corresponding to the memory's size) and

[1] The code is available at https://github.com/UWCCDL/PTSD

the number of attributes from which a value for a slot can be chosen from.

Although certain PTEs do occur in familiar environments and conditions (i.e., domestic abuse), many traumatic memories have origin in situations, conditios, and environments that are unique and different from the daily life of the agent (i.e., war). The degree of exceptionality of the PTE was captured by parametrically variying the pool of attributes that were chosen for the PTE's slots. Specifically, a parameter $0 \leq M \leq 1$ controlled the proportion of attributes that were unique to the PTE, that is, were selected from a special pool instead of being drawn from the same pool as the attributes of other memories. For $M = 0$, the PTE is entirely made of unique attributes, while for $M = 1$ the PTE is absolutely indistinguishable from the other randomly generated memories.

Implementation of Emotion Effects

To simulate the effects of emotion and trauma on declarative memory, the ACT-R code was augmented with a set of functions that manage a new dictionary $\mathcal{V} : i \rightarrow V(i)$ that maintains the scalar value $V(i)$ for every memory chunk i. Every time a new chunk is created, its value $V(i)$ is computed and added to $\mathcal{V}$ (see Algorithm 1).

The value of $V(i)$ is determined on the bases of the time t at which i is first added to ACT-R's declarative memory system. If $t = t_{PTE}$, then the value of $V(i)$ is set to $V(i) = V_{PTE}$. Otherwise, the value is drawn from a uniform distribution with limits $[0,2]$.

ACT-R's baseline equation (Eq. 8) is then augmented with the new term corresponding to $\log V(i)$, which is calculated by taking the logarithm of the entry of i in $\mathcal{V}$. Because values $V(i)$ for non-traumatic events are uniformly distributed between 0 and 2, $overline V = 1$ and $\log \overline{V} = 0$, so the bias term in Eq. 10 can be ignored.

Spreading Activation

In ACT-R, spreading activation depends on the links between contextual cues and the memory being retrieved. Since, in our model, all of the contextual cues $q_1, q_2 ... q_N$ are embedded into a single chunk, the values of $s_{q,i}$ were simplified as follows: if two chunks share the same attribute in the same position, then $s_{j,i} = 1/N$, otherwise $sj, i = 0$. This ensures that $0 \leq N \sum_q s_q \leq 1$.

Algorithm 1: Simulation Loop

Input: Traumatic Value V_{PTE}, Simulation time T

while $T \leq T_{End}$ **do**
- $C \leftarrow$ `NewChunk`(N_S, N_A);
- **if** $T = T_{PTE}$ **then**
  - $\mathcal{V}[C] \leftarrow \mathrm{V}_{PTE}$;
- **else**
  - $\mathcal{V}[C] \leftarrow$ `Uniform`$(0,2)$
- Imaginal $\leftarrow C$;
- Retrieval $\leftarrow i : \mathrm{argmax}(A_i)$;
- T $\leftarrow T + \Delta T$

Simulations

Simulations were run by systematically varying the model parameters as shown in Table . For each combination of parameters in parameter space, the model was run for 100 times.

Table 1: Sample table title.

Parameter	Value (or Range)
T	300,000 s
ΔT	1,200 s
T_{PTE}	18,000 s
N_S	6
N_A	6
M	0, 0.25, 0.5, 0.75, 1
W (Eq. 7)	0, 2.5, 5, 7.5, 10
d (Eq. 6)	0.1, 0.3, 0.5, 0.7, 0.9
V_{PTE}	1,5, 10, 15, 20

Results

The occurrence of intrusive memories was measured as the probability of retrieving the PTE at any retrieval cycle. For simplicity and ease of interpretation, the results are shown in terms of probability of retrieval per simulated day, with the entire simulation (T) extending to approximately two months after the occurrence of the PTE.

A complete analysis of the model simulations can be found on the project repository [2]. Here, we will overview the main findings.

Recovery Trajectories

In general, at different parameters the model exhibits different recovery curves following the PTE. For reasonable values of $d > 0.5$ and $W > 5$, the model seems to produce the four major trjaectories identified by Bonnano (Bonanno & Mancini, 2012).

A number of effects are consistent with the classic results in the literature. The model's recovery trajectory, as expected, generally worsens with the intensity of the traumatic event (V_{PTE}). The model's trajectory also improves with greater executive function and working memory (W). Finally, traumatic memories occurring in a familiar environment ($M > 0.5$) are more difficult to suppress, and more likely to become intrusive, than those occurring in exceptional circumstances ($M < 0.5$).

As expected, the model's trajectory often exhibit *chaotic behaviors*, with small changes in one parameter often resulting in dramatic changes in the recovery curves (technically, *bifurcations*). For example, for $W = 7.5$ and $d = 0.7$, and increase in similarity from $M = 0$ to $M = 0.25$ transforms

[2] https://github.com/UWCCDL/PTSD/ptsdanalysis.html

a "spontaneous" recovery curve (with the probability of an intrusive memory declining over time) into a "chronic" one (with the probability catastrophically increasing over time).

Acknowledgments

This research was supported by a scholarship from the University of Washington's Institute for Neuroengineering (UWIN) to the first author.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum Associates.

Anderson, J. R. (2009). *How can the human mind occur in the physical universe?* (Vol. 3). Oxford University Press.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Bonanno, G. A., & Mancini, A. D. (2012). Beyond resilience and ptsd: Mapping the heterogeneity of responses to potential trauma. *Psychological trauma: Theory, research, practice, and policy*, *4*(1), 74.

Fum, D., & Stocco, A. (2004). Memory, emotion, and rationality: An act-r interpretation for gambling task results. In *Iccm* (pp. 106–111).

Juvina, I., Larue, O., & Hough, A. (2018). Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making. *Cognitive Systems Research*, *48*, 4–24.