Modeling Memory Dynamics in Post Traumatic Stress Disorder: Intrusive Memories, Recovery Trajectories, Neurobiological Effects, and Fear Overgeneralization

Briana Marie Smith

A these

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2023

Reading Committee:

Azadeh Yazdan-Shahmorad, Chair

Andrea Stocco

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Modeling Memory Dynamics in Post Traumatic Stress Disorder: Intrusive Memories, Recovery Trajectories, Neurobiological Effects, and Fear Overgeneralization

Briana Marie Smith

Chair of the Supervisory Committee:
Washington Research Foundation Innovation Assistant Professor Azadeh Yazdan-Shahmorad
Bioengineering and Electrical Engineering

Post-traumatic stress disorder (PTSD) is a psychiatric disorder often characterized by the unwanted re-experiencing of a traumatic event through nightmares, flashbacks, and/or intrusive memories. This thesis presents a neurocomputational model using the Adaptive control of Thought - Rational (ACT-R) cognitive architecture that simulates intrusive memory retrieval following a potentially traumatic event (PTE) and derives predictions about an individual's recovery trajectory, neurological effects including changes in hippocampal volume and functional connectivity, and behavioral symptoms. Memory intrusions were captured in the ACT-R framework by weighting the prior probability of re-encoding a memory by an emotional *intensity* term *I*, which captures the degree to which an event was perceived as dangerous or

traumatic. A series of simulations were run in which the model performed memory retrieval under naturalistic conditions for up to two months after experiencing a simulated PTE. It was found that *I* had a significant effect on the probability of experiencing traumatic memory intrusions following a PTE, and that, under different conditions, the model experienced different probabilities of undergoing different recovery trajectories. The model also found that *I* was a significant predictor of hippocampal volume reduction, where the mean and range of simulated volume loss match results of existing meta-analyses. Functional connectivity between the medial prefrontal cortex and hippocampus was shown to be significantly greater in the simulated control group than in the simulated PTSD group. Finally, the model accounts for the behavioral effect of fear overgeneralization in PTSD, predicting greater startle probability to novel compound stimuli after a fear conditioning task in simulated PTSD groups compared to a simulated control group. The authors believe that this is the first model to describe traumatic memory retrieval, predict recovery trajectories, and provide a mechanistic account of multiple symptoms of PTSD including intrusive memories, neurobiological changes in hippocampal volume and functional connectivity, and fear overgeneralization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To my advisors Azadeh Yazdan-Shahmorad and Andrea Stocco. Your mentorship has helped me grow professionally and personally in so many ways—thank you. To my co-authors and collaborators, Emma PeConga, Patrick Rice, Catherine Sibert, Madison Thomasson, Cher Yang, and Lori Zoellner. Your contributions have strengthened this work and allow it to continue growing in impact and effectiveness. Finally, to my fellow graduate students, classmates, and friends for your support, contributions, and friendship as this work developed from an idea to a model, and beyond.

DEDICATION

To my family, for your love, encouragement, and support as I made the journey through graduate school and in all the challenges we've faced along the way. To my mom, for inspiring this work and for doing the best she could. She gave me a better life than she was given, and because I can't tell her how grateful I am, I will say it here. Thank you, mom. You were so strong, and I love you. Finally, to the new family I've found in Al-Anon. The road to recovery is not easy and you have given me love, support, and hope.

Chapter 1. INTRODUCTION

Intrusive memories are emotionally charged memories of traumatic events whose unwanted recollection disrupts normal function and causes additional stress. They are a transdiagnostic symptom across different mental health conditions, such as depression, schizophrenia, social anxiety disorder, and generalized anxiety disorder [1]. The timecourse of intrusive memories is important for understanding and predicting recovery trajectories, which remain highly variable and difficult to characterize in clinical practice and research.

Perhaps the most prototypical pathology characterized by intrusive memories is post-traumatic stress disorder (PTSD), a condition of severe distress triggered in response to a traumatic event. PTSD offers a model situation to study intrusive memories. The onset of PTSD is typically tied to a specific event in time (the potentially traumatic event, or PTE) and intrusive re-experiencing of the event is an important part of its related symptomatology. Intrusive memories are distinguished from the normative re-experience of events because of their unwanted and uncontrolled nature, increased frequency, persistence over time, and the severity of distress that is associated with them [2].

The goal of this project is to use the tools of computational cognitive neuroscience to provide a mechanistic explanation of how intrusive memories arise and which factors affect the possible recovery trajectories. Specifically, this thesis presents a computational model that describes a potential mechanism by which intrusive memories arise and makes predictions about their frequency over time, their possible recovery trajectories, and their biological and behavioral consequences. A central assumption that sets this approach apart from previous models [3] is that traumatic memory intrusions can be understood within the context of a general theory of episodic

memory. This theory is grounded in a rational analysis of memory, is well specified computationally, and has a clear interpretation in terms of biology, thus providing a solid conceptual framework to understand and model traumatic memories.

This approach might seem paradoxical because, while most models of episodic memories stress its Bayesian rationality and the adaptiveness of remembering and forgetting processes [4], intrusive memories are considered maladaptive, dysfunctional, and disruptive to a normal life. As the reminder of this thesis will show, this paradox might be illusory: intrusive memories can be modeled following the same adaptive laws that regulate normal memories; however, the exceptional conditions under which traumatic memories are encoded lead the normally self-regulating mechanisms of forgetting to out-of-the-ordinary, self-perpetuating dynamics. Within this framework, the persistent memory intrusions observed in PTSD can be seen not as a maladaptive response, but rather as the runaway process of an otherwise adaptive memory system.

A traumatic memory tends to out-compete more contextually appropriate memories due to the fact it was encoded in a highly emotional state. Traumatic experiences evoke a physiological response that is accompanied by increased activation of subcortical areas such as the amygdala [5,6]. Intrusive memories are thought to occur because of the simultaneous activation of the amygdala and hippocampus during memory encoding [7]. With each retrieval of the traumatic memory, disproportionately more resources are allocated to it, leading to the further preservation and growth of these unwanted memory intrusions. Under different initial conditions, this positive feedback loop might not be initiated or, if it is initiated, might extinguish itself, stabilize, or worsen, leading to different recovery trajectories. Within this framework, it is proposed that the neurobiological changes following a traumatic event are the result of different

computational demands posed to the biological memory circuitry. Specifically, potential changes in hippocampal volume associated with PTSD [8] can be explained as the natural result of allocating neural resources to changing memory demands, and potential changes in the functional connectivity between the frontal cortex and the hippocampus [5] can be explained as an increased bottom-up, and reduced top-down, control over the contents of memories being retrieved.

Finally, it is important to note that traumatic experiences are not the only instance when enhanced memory encoding is observed. It is normal and adaptive for a spectrum of experiences to elicit physiological or emotional responses that increase brain activation and memory encoding. The memory encoding effects of both positive and negative experiences can be quantified by measuring the degree to which the emotional state is generalized onto a neutral stimulus. Pavlovian conditioning is a well-known paradigm that demonstrates the positive effect of emotion on memory. Fear conditioning is similar but in the negative valence, where the emotion of fear, in the form of a startle or freeze response, generalizes to neutral stimuli. The startle reflex is a cross-species response to loud noises, sudden movement, sharp pain, etc., and the generalization of these memories to other environmental cues is an adaptive neurobiological process that promotes survival in complex and dynamic environments [9]. A characteristic of PTSD is impaired fear inhibition and the tendency to overgeneralize an otherwise normal fear response to neutral stimuli more often than healthy controls. In this thesis, we will show that fear conditioning and fear generalization arise as a natural consequence of the rational Bayesian framework and that the same framework produces a natural account for how previous traumatic experiences modify learning in later contexts.

The remainder of this thesis is structured as follows. First, we will review the main findings on intrusive memories in PTSD, including some critical factors that are known to affect them. Second, we will introduce the model and provide an interpretation of its internal parameters in terms of cognitive constructs and neurobiology. Then, we will examine the model's performance across a series of parameter values and examine the degree to which it correctly replicates the main findings in the literature in terms of behavioral and neurobiological findings. Finally, we will review the model, its limitations, and possible directions in which it could be refined.

1.1 Intrusive Memories in PTSD

Although intrusive memories appear as symptoms in multiple psychological disorders, this thesis will focus on modeling their occurrence in PTSD. PTSD is usually understood as a persistent pathological response to an external event, the potentially traumatic event (PTE). The existence of such an external event provides a natural way to anchor the evolution of symptoms from a particular point in time and makes PTSD a convenient starting point for any modeling effort that ties together memory processes and psychopathology.

In examining intrusive memories, two important outcomes need to be considered. One is the absolute outcome at a time point, which can be measured, for example, by the number of daily or weekly intrusive memories experienced by a patient. Their frequency, together with the characteristics of their context and the context in which they occur, are highly correlated with the severity of PTSD symptoms, explaining up to 43% of the variance in severity six months after the PTE [10].

The second outcome of interest is the patient's recovery trajectory, i.e., whether the patient is on their way to improving or not. In a recent and influential review, Galatzer-Levy et al. [11]

identified four typical recovery trajectories following the PTE, which are illustrated in Figure 1.1. In this figure, the x-axis represents time, the vertical black line represents the onset of a PTE, and the y-axis represents the severity of intrusive re-experiencing on a qualitative scale. After a PTE, an individual might not show intrusive re-experiencing symptoms (resilient trajectory, green); show acute symptoms that spontaneously decline in severity over time (recovery trajectory, blue); exhibit worsening symptoms over time (delayed trajectory); or never recover after showing an initial acute response (chronic trajectory).

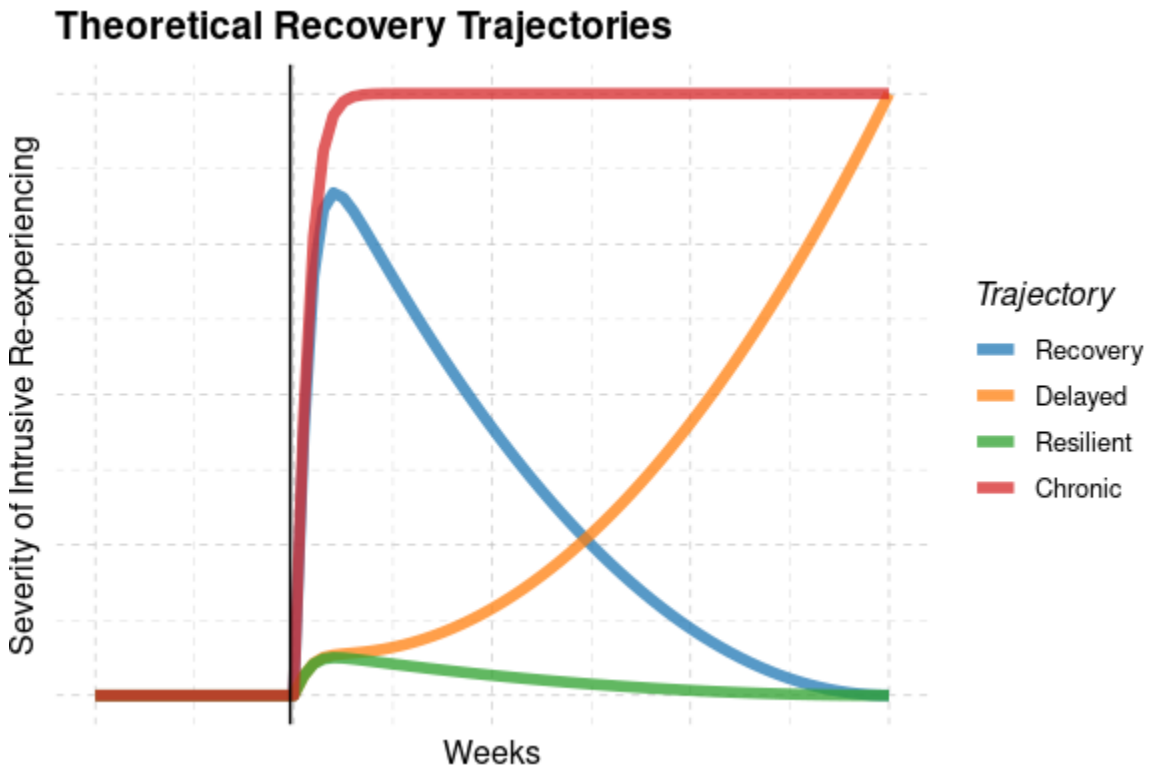

Figure 1.1. Idealized recovery trajectories after a PTE (black vertical line).

Because these trajectories were derived from an overview of studies that differ greatly from each other, they are idealized in two important ways. First, all of the non-resilient trajectories appear to reach the same maximum level of symptoms on the y-axis, while this does not need to

be the case. Second, an individual's trajectory might be categorized differently depending on the time scale at which it is observed. Different studies, for example, might examine clinical outcomes weeks, months, or years after the PTE; it is possible that an individual who is categorized as chronic two months after the PTE would be considered recovering one year afterwards, and that an individual exhibiting a resilient phenotype after one month could experience a delayed onset at a longer time scale.

These limitations notwithstanding, the existence for these different trajectories is generally strong, as variants of these trajectories can be found across a variety of studies and reviews, and at least three of them can be identified in animal model studies (evidence for the delayed onset trajectory is debated, and this trajectory does not typically appear in animal studies of fear conditioning) [12]. Importantly, the existence of different trajectories poses significant challenges for the early identification of those at high risk for chronic psychopathology—an important clinical goal not only for preventing long-term suffering but also for appropriately allocating limited resources. For this reason, it is important to characterize and understand the nature of the factors that might lead to these different patterns.

While some of the factors could be attributed to environmental effects, at least some of the differences in recovery curves must be due to the idiographic characteristics of the patients. Evidence in this sense comes from a review of studies on fear generalization and extinction in rodents (a common paradigm to study the neurobiology of PTSD), which concluded that at least three of the most common trajectories (chronic, spontaneous recovery, and resilience) also appear in animal models of PTSD [11].

1.2 Factors Affecting the Frequency of Intrusive Memories Following a PTE

In a recent review, Marks et al. [7] provide an exhaustive examination of the possible factors that have been associated with the frequency and severity of intrusive memories. Many factors that had been suggested did not show more than modest effects in Mark's review. For example, trait dissociation did not consistently or robustly predict intrusive memories in most studies. Some of the factors that emerged as having significant effects, on the other hand, present significant difficulties for a modeling approach. For example, the presence of trait anxiety and depression were a small to moderate predictor of PTSD; yet, no consensus exists on how these pathologies should be modeled, as the field of computational psychiatry is still in its infancy (in fact, the present study is an attempt to advance this very field). Similarly, although gender is an important factor determining both the severity and the type of PTSD symptoms [13], it cannot be easily captured through model parameters. Finally, some factors that are robust predictors but are, in fact, correlated with other factors. For example, the tendency towards "data-driven" processing, defined as encoding sensory-perceptual details without the broader conceptual context of the event and typically assessed via self-report, does predict intrusive memories but likely occurs more in those with trait anxiety and increases as state anxiety increases.

Thus, the modeling efforts described herein focused on a handful of factors that (1) stand out as having a robust and well-documented effect, (2) can be naturally interpreted within our modeling framework, and (3) are representative of different categories. For simplicity, we will divide these factors into three categories: environmental factors, idiographic factors, and the tendency for negative re-appraisals, which is often tied to the presence of other conditions like anxiety and depression (Table 1.1).

Table 1.1. Factors affecting the frequency of intrusive memories in PTSD

Category	**Factor**	**Relationship to Mental Health Outcome**	**Model Parameter**
Environmental	Intensity of the traumatic event	↑Greater intensity = ↓worse outcome	*I*
	Similarity between the PTE and daily life events	↑Greater similarity = ↓worse outcome	*C*
Idiographic	Cognitive control	↑Greater control = ↑better outcomes	*W*
	Vividness of recollection	↑Greater vividness = ↓worse outcomes	γ
Rumination or Negative Appraisal	Unproductive Processing	↑Greater unproductive processing = ↓worse outcome	*U*

1.2.1 *Environmental Factors*

We use the term "environmental factors" to indicate factors related to the characteristics of the PTE and their relationship to an individual. Two environmental factors were selected. The first is the emotional intensity of the traumatic event. The PTSD literature covers a vast variety of traumatic experiences, spanning the gamut from witnessing a crime to surviving torture. Although all of these events are traumatic, they vary in degree and severity, and these variations are reflected in their effects on the individuals affected by them. Traumatic events of greater perceived intensity tend to be associated with the worst outcomes [13], often in a dose dependent manner [14].

Traumatic events vary not only in their intensity, but also in the circumstances in which they might occur. These circumstances, and specifically the degree to which they are perceived to overlap with the daily routines, also profoundly affect the severity of PTSD symptoms. Sexual assault victims, for example, might have experienced trauma in their very own workplace or home, while the trauma experienced by combat veterans typically occurs in situations that have little overlap with their civilian life. The similarity between the circumstances of the PTE and the

context of everyday life plays an important role in triggering intrusive memories, since a more similar environment offers more contextual cues for retrieval [15].

1.2.2 *Idiographic Factors*

It is generally understood that differences in outcomes are an interaction of environmental and idiographic factors, that is, factors that capture the specific cognitive traits of a patient. This thesis focuses on two such factors. The first such factor is a patient's capacity to exert control over their own thoughts and habitual responses. This capacity is known as cognitive control or executive function and it is often referred to in the clinical literature [16–18]. Many PTSD studies have also focused on related psychological constructs, such as response inhibition [19,20], working memory [21,22] and intelligence [23]; all of these concepts are highly overlapping [24]. However operationalized, the existing clinical research shows that individuals with better capacity for cognitive control experience less or less-severe PTSD symptoms, likely because of a better ability to manage or inhibit unwanted intrusive memories.

Another important factor is the vividness of the recollection experience, that is, the degree to which a retrieved memory carries the same emotional intensity as the original event. This factor also appears, under different studies, with different names. Some studies discuss it in terms of sensorimotor vividness, that is, greater detail and imagery in the re-experience of sensory details of the traumatic event [25,26]. Other studies discuss it in terms of emotional vividness, typically in terms of the emotional intensity of the memory [27]. In general, the vividness of the details of the memory is highly associated with its emotional intensity, so the two are considered the same herein.

1.2.3 *Unproductive Processing*

As noted in the previous sections, intrusive memories are trans-diagnostic and occur in multiple disorders. Conversely, the co-occurrence of other disorders represents a risk factor for PTSD, exacerbating the tendency to retrieve unwanted traumatic memories. This tendency to further retrieve and elaborate on negative thoughts and memories is categorized as either *rumination* or *negative appraisal*. Rumination is the tendency to perseverate on negative thoughts and feelings, whereas negative appraisal is the tendency to reflect negatively on previous events. Not surprisingly, for individuals whose response style is characterized by ruminative processing or negative appraisals of their traumatic memory, the clinical outcome is usually worse [7].

Although there are important differences between these two clinical concepts, they share the common nature of conscious and deliberate retrieval of memories, among which the traumatic ones are likely to figure prominently. Such a process has a clear connection to the mechanics of the model proposed here: by increasing the frequency with which a traumatic memory can be retrieved, both rumination and negative appraisal contribute to enhancing its availability and, therefore, the future probability of intrusive memories. For simplicity, the model proposed herein does not distinguish between rumination and negative appraisal, and we will refer to this common phenomenon simply as *unproductive processing*.

1.3 Neurobiological Effects of PTSD

In addition to psychological symptoms, trauma may also leave recognizable traces in the survivors' brain activity. A prevalent view in the literature organizes these changes in the context of the so-called Triple Network model, a framework that highlights changes in three interrelated

networks of functionally and anatomically connected brain regions that are involved in controlling emotion and memory [28]. The first of these networks is the so-called salience network, which involves brain regions associated with quickly detecting the emotional relevance of external stimuli, including the insular cortex, the anterior cingulate cortex and, most notably, the amygdala. Hyperactivity in this network, and in particular in the amygdala, is associated with PTSD symptoms [29]. For instance, compared to trauma-exposed controls, patients with PTSD symptoms show increased amygdala activity to negative stimuli [5,30,31].

The second network is the Default Mode Network (DMN), a large network of regions comprising the medial prefrontal regions, the medial parietal and retrosplenial cortex, and, crucially, the hippocampus. The hippocampus is necessary for long-term memory formation and storage of episodic and autobiographical memories [32–34], although some researchers suggest that episodic memories might eventually be entirely transferred to neocortical sites [35,36]. In addition to the hippocampus, other regions in the DMN are involved in many aspects of episodic memory encoding and retrieval [37].

Intrusive memories are thought to occur because of the simultaneous engagement of the amygdala and hippocampus at the time of memory encoding, with amygdala re-activation later driving the spontaneous and recovery of the original traumatic event in the hippocampus [7]. These spontaneous intrusions occur outside the scope of the usual mechanisms that control memory retrieval processes, which are associated with the third network: the central-executive network (CEN). This network spans dorsal prefrontal and superior parietal regions that are involved in higher-level control and coordination. Consistent with the presence of uncontrolled or involuntary memory intrusions, individuals suffering from PTSD exhibit reduced activity in regions belonging to this network, such as the medial prefrontal cortex [5].

Although the majority of studies have identified PTSD-specific changes in brain activity during tasks, the facets that are potentially more relevant to this thesis are the long-term, stable changes that manifest themselves independently of the task, such as the possible changes in brain structure and in a spontaneous network activity.

At the anatomical level, PTSD is also generally characterized by a marked reduction in the volume of the hippocampus [8,38–42]. It is important to note that this change is primarily structural, and, although often remarkably apparent, decreased hippocampus size is not consistently accompanied by a functional impairment in long-term memory performance [43]. Nonetheless, reductions in hippocampal volume in the range of 5-26% have been observed across multiple studies [8]. Only a few studies have been able to examine hippocampal volume prior to trauma exposure and PTSD [44,45]. When looking cross-sectionally, the majority of studies suggest that changes in hippocampal volume may be a neurological manifestation of PTSD. The model reported herein provides an explanation for this finding, relating the change in structure to the differential neural resources required to accommodate intrusive memories.

A second important change occurs in the resting-state functional connectivity between different regions in the default mode network. Resting-state functional connectivity is measured as the correlation in low frequency (0.1 - 0.01 Hz) oscillations in spontaneous activity in the brain, which are recorded while participants are awake, but not actively doing any particular task, in an fMRI scanner. Thus measured, functional connectivity is stable and replicable, and assumed to index the spontaneous architecture of brain networks. Compared to healthy or traumatized, asymptomatic controls, individuals with PTSD show decreased functional connectivity between the medial prefrontal cortex and the hippocampus. This change is believed

to reflect a reduced capacity of top-down, goal-oriented control on the contents of memories that are being retrieved [5,29].

1.4 Behavioral Effects of PTSD: Fear Inhibition

Fear conditioning is a behavioral aversive learning paradigm that is commonly used to study memory and learning in animals and humans under controlled, laboratory conditions. In a fear conditioning task, a neutral conditioned stimulus (CS) is paired with an aversive unconditioned stimulus (+), such as an airblast or electric shock. The *CS+* stimulus combination is presented repeatedly in order to create an association between the *CS* and + such that later presentation of *CS* alone is sufficient to elicit an observable fear potentiated startle response. The startle reflex is a cross-species response that makes fear conditioning an experimental model that bridges the gap between human and animal research on stress and anxiety [46].

The primary mechanism governing this fear conditioning response is that the generalization of fear memories is an adaptive neurobiological process that promotes survival in complex and dynamic environments [9]. It has been established in the human literature however, that individuals suffering from PTSD are more sensitive than controls to the fear conditioning paradigm, producing a higher percentage of startling responses for both conditioned and unconditioned stimulus pairs [47,48]. This fact is typically taken as strong evidence of reduced fear inhibition in which individuals primed by traumatic memories generalize fear to a higher extent than controls, contributing to the theory that previous unrelated trauma enhances future aversive learning [47–49].

1.5 Intrusive Memories and Existing Models of PTSD

As noted above, the biological mechanisms of PTSD are typically understood through the lens of fear conditioning and extinction literature. In computational models, conditioning and extinction are captured through Reinforcement Learning [50], a formal theory of how agents learn to select different actions based on environmental feedback [3,51–54]. For example, Myers et al. [52] developed an RL model with separate mechanisms for rewards and losses, and showed that performance of PTSD participants in a probabilistic decision-making task, could be modeled as a consequence of abnormally large reward values ascribed to neutral trials. A related model [53] showed similar results in groups of participants with the personality trait of behavioral inhibition, typically marked by higher anxiety and avoidance behavior. Individuals with high behavioral inhibition showed higher reward values assigned to neutral faces in a trust task than individuals with lower inhibition. Both findings can be interpreted as related to negative expectations connected to avoidance and impaired performance in learning from rewards, both of which are potentially implicated in PTSD.

As noted by Radell et al. [3], these models are rarely designed to capture individual differences. They are also geared towards capturing laboratory-based paradigms rather than naturalistic circumstances, and task-based performance measures rather than clinical outcomes such as recovery trajectories or estimates of clinical symptoms. To the best of our knowledge, none of these existing models focuses specifically on intrusive memories nor captures long-term memory dynamics and clinical symptoms. These dynamics, instead, were explicitly included in an influential framework proposed by Rubin et al. [55]. Rubin's theoretical model shares the same viewpoint as the model presented in this thesis, but lacks a computational implementation and, therefore, yields only qualitative predictions. Thus, the model proposed herein offers a new

perspective to the study of intrusive memories in PTSD from a computational neuroscience viewpoint. Recently, and in parallel with the research described herein, Cohen and Kahana have presented a computational model of PTSD based on abnormal cue retrieval [56]. Like my model, Cohen and Kahana based their analysis on the dynamics of episodic memory. Unlike this proposal, however, their model does not include explicit mechanisms of forgetting and decay, nor does it include a general theory of how and why emotion should affect memory.

1.6 Summary and Goals

The brief review of the literature outlined above clearly points to the reasons why a new model is needed. Effective mental health preventative care and treatment measures depend on the ability to successfully predict recovery trajectories, but the ability to predict recovery trajectories is hindered by the existence of multiple factors that might affect the outcome. To further complicate the problem, these factors are often difficult to measure objectively and their relationship with the outcome might be indirect and non-linear.

Computational models are a theory-driven approach to capture the effects of the various factors and constrain their inter-dependencies into a principled set of relationships and assumptions [57,58]. In addition, if computational models are rooted in cognitive and neurobiological frameworks, their internal parameters naturally reflect established and measurable constructs, facilitating their interpretation of their effects on patient mental health [59,60]. Finally, a computational model might provide a principled way to quickly test and revise assumptions in developing future, better-targeted PTSD prevention interventions and treatments.

The next section will provide an overview of the proposed model, explaining how the model relates to an existing model of episodic memory; how these models can be extended to

account for the specific characteristics of PTSD; and how the parameters governing the model relate to concepts rooted in both the neurocognitive and clinical literature.

Chapter 2. A COMPUTATIONAL FRAMEWORK FOR TRAUMATIC MEMORIES

2.1 THE RATIONAL ANALYSIS FRAMEWORK

The model presented herein is based on the Adaptive Control of Thought – Rational (ACT-R) theory of declarative memory [61]. This choice was motivated by three reasons. First, ACT-R is currently the most commonly adopted cognitive architecture in psychology and cognitive neuroscience [62]. Second, ACT-R has a long and established history of application to brain sciences, making the process of drawing new inferences at the neural level easier and less tentative. Finally, ACT-R is based on a Bayesian framework, which provides an elegant foundation of declarative memory and can be extended to incorporate the proposed theory of memory retrieval that includes emotional intensity.

Despite previous attempts [63–65], ACT-R currently lacks a formal theory of emotion. Thus, it is not possible to implement the effects that traumatic experiences have on memory without extending the underlying theory. To provide a rational justification for such an extension, we will follow the Bayesian interpretation of memory that was the original inspiration for ACT-R's memory system [66].

In the rational analysis terms pioneered by Anderson [4], a memory's retention function (that is, its availability across contexts and times) reflects its probability of being needed at a given moment. Thus, if we indicate the specific memory as m and the context as Q (composed of

different elemental cues $q_1, q_2, \ldots q_n$), a memory retention function reflects the posterior odds $P(m|Q) / P(\neg m|Q)$, which can be expressed as the product of the prior odds $P(m) / P(\neg m)$ and the likelihood $P(Q|m) / P(\neg Q|m)$. Assuming, for simplicity, that each cue q is independent from each other, the retention function can be expressed in terms of log odds as:

$$\begin{aligned}\log\left(\frac{P(m|Q)}{P(\neg m|Q)}\right) &= \log\left(\frac{P(m)}{P(\neg m)} \times \frac{P(Q|m)}{P(Q|\neg m)}\right) \\ &= \log\left(\frac{P(m)}{P(\neg m)}\right) + \log\left(\frac{P(Q|m)}{P(Q|\neg m)}\right) \\ &= \log\left(\frac{P(m)}{P(\neg m)}\right) + \log\left(\prod_q \frac{P(q|m)}{P(q|\neg m)}\right) \\ &= \log\left(\frac{P(m)}{P(\neg m)}\right) + \sum_q \log\left(\frac{P(q|m)}{P(q|\neg m)}\right) \\ &= \log\left(\frac{P(m)}{P(\neg m)}\right) + \sum_q \log\left(\frac{P(q|m)}{P(q)}\right) \qquad (2.1)\end{aligned}$$

The last step in Eq. (2.1)(2.1) is an approximation derived from the consideration that, for large amounts of memories, $P(q \mid \neg m) \approx P(q)$. The terms in Eq. (2.1)(2.1) have a straightforward explanation in terms of the cognitive psychology of memory [4,61,67]. Specifically, the log posterior odds on the left-hand side of Eq. (2.1)(2.1) correspond to a memory's activation, *A(m)*, an intuitive construct that describes a memory's moment-to-moment availability. Similarly, the two quantities on the right-hand side also correspond to cognitive constructs, with the log priors corresponding to the *base-level activation* of *m* or $B(m)$ (capturing the effects of the previous history of usage of *m*) and the log-likelihood corresponding to the *contextual* or *spreading activation* of *m*, or $S(m)$ (capturing the additive effects that each environmental cue has on the memory's activation). Thus:

$$A(m) = \log\left(\frac{P(m|Q)}{P(\neg m|Q)}\right)$$

$$B(m) = \log\left(\frac{P(m)}{P(\neg m)}\right)$$

$$S(m) = \sum_q \log\left(\frac{P(q|m)}{P(q)}\right) \quad (2.2)$$

By combining Eq. (2.1) and (2.2), we get the simplified activation equation:

$$A(m) = B(m) + S(m) \quad (2.3)$$

To implement these equations algorithmically [68], the quantities $B(m)$ and $S(m)$ are approximated in ways that predict future use based on the previous history or the learned associations between the memory and the contextual cues, respectively.

In general, the probability of retrieving a memory declines over time according to a power function [66]. Assuming that every single use of a memory *m* corresponds to the creation of a new trace [32] for *m*, and that the probability of each trace being retrieved declines over time according to the same power function with a decay rate *d*, then the odds of retrieving *m* are the odds of retrieving *any* of the traces associated with *m*, or the sum of the probabilities of each trace being retrieved. Because activation is expressed in the form of log odds, the value of $B(m)$ at time *t* is the log of the sum of decaying traces associated with *m*'s usage:

$$B(m) = \log\left(\sum_j (t - t_j)^{-d}\right) \quad (2.4)$$

in which t_j represents the time at which trace *j* has been encoded, and $t - t_j$ is the age of the trace. Figure 2.1 provides a visual illustration of this mechanism, assuming the same memory has four traces associated with it, generated at times $t_1 = 0$ (memory creation).

The model is noteworthy for its reliability, having been used to successfully model effects of recency, frequency, and spacing in declarative learning [69–71] and having been used to successfully derive optimal schedules for learning practice [72]. The decay parameter *d* in Eq. (2.4) has been also used for deriving idiographic parameters [73], with *d* being a stable trait within the same individual across sessions and materials, and to assess individual differences in

real-life outcomes, such a student's success at answering test questions after studying [74]. Notably, these studies have shown that the same model and the same parameter values can be used to capture effects at different time scales, from minutes in laboratory settings [69] to weeks and months in naturalistic tasks [74].

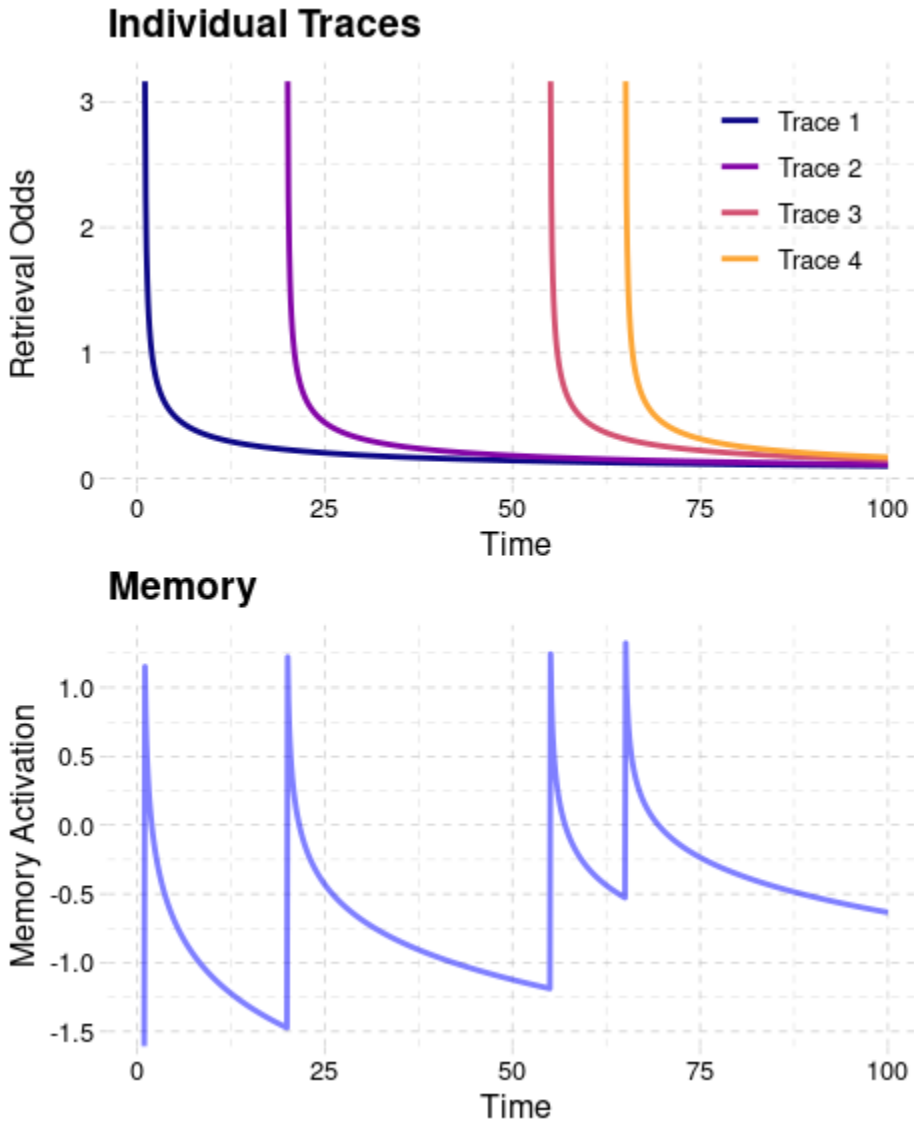

Figure 2.1. Time course of base-level activation of a memory m which has four traces associated with it, encoded at times t = 0, 20, 55, and 65 s (top). Each trace, shown in a different color, decays

at the same rate (d = 0.5). The odds of retrieving m are the sum of the odds of retrieving any of its traces (bottom).

The base-level activation component provides a first clue as to the reason why intrusive memories exist. As shown in Figure 2.1, the more a memory has been retrieved, the more of its traces are created and the more likely it is to be retrieved in the future. Intrusive memories feed off this frequency effect because more unwanted retrievals in the past lead to more unwanted retrievals in the future. Under normal circumstances, this cycle is broken by the natural decay of each trace (the recency effect). As it will be shown, the special circumstances under which traumatic memories are created makes them more likely to become persistently intrusive. The effect of previous retrievals is also related to one of the factors affecting the frequency of intrusive memories, namely, the effect of negative appraisal. This is because a tendency to revisit negative events would likely result in more retrievals of a traumatic memory, ultimately increasing its odds of intruding in the future.

Spreading activation $S(m)$, on the other hand, can be best understood in reference to a classic representation format for memories, namely, semantic networks [75–78]. In semantic networks, each memory represents a node, and associated memories. are connected by directional links. Spreading activation $S(m)$ is implemented as the amount of activation that flows from the memory nodes that are part of the context $Q = \{q_1, q_2 \ldots q_N\}$. Thus, each element q_j represents a specific active node in the current context. The strength of the link between q and m reflects the statistics of co-occurrence between the two events, so that greater co-occurrence of q when m is present (i.e., $P(q|m)$) corresponds to stronger links. If there is a direct link from q to memory m, then m receives an activation boost that is proportional to the product between the strength of the link connecting q to m (indicated as $S_{q \rightarrow m}$) and the attentional weight given that

cue. The attentional weight is usually simplified as a single scalar quantity, *W*, that is divided by the number *N* of features that are present in the context *Q*. The total amount of spreading activation *S(m)* that *m* receives is the sum of all of the partial effects of each element *q*:

$$S(m) = \sum_q \frac{W}{N} S_{q \to m} \quad (2.5)$$

The idea that spreading activation is related to attention has a long tradition in cognitive psychology [75,76,78]. Although the most common interpretation of the weight parameter *W* is in terms of top-down attentional control in retrieval [75], this parameter can also be interpreted as capturing executive functions and *working memory* [79–81], that is, an individual capacity to maintain, process, and update short-term information [82,83]. For example, Lovett and colleagues [84,85] showed that individual variations in *W* values capture idiographic differences in working memory performances, and that *W*, when estimated independently through a working memory task, predicts performance on other tasks that demand cognitive control. Similarly, Rice and Stocco [86] showed that individual variations in performances in the *N*-back task (a common measure of working memory) can be explained by corresponding differences in the *W* parameter of the corresponding model.

Thus described, spreading activation is related to two important factors that affect PTSD. The first is the contextual similarity. In this framework, the degree of similarity between the context in which the PTE occurred and the context of daily activities is captured by the amount of individual cues *q* that are shared between the two; the greater the number of shared cues, the greater is the spreading activation and thus the probability of involuntarily retrieving a memory related to the PTE. The second factor is executive function. In this framework, greater executive function can be translated into greater values of the parameter *W*. The greater is *W*, the greater is the amount of top-down control exerted over the retrieval process and, correspondingly, the

smaller is the probability of involuntarily retrieving a traumatic memory in a different context. Although spreading activation provides a way to capture these factors, two more factors (emotional intensity and recollection vividness) still need to be accounted for. To do so, the model will need to be expanded to account for the effects of emotion on memory.

2.2 Extending the Rational Analysis Framework

It has been noted several times, even by Anderson himself [61], that one limitation of the classic approach to memory summarized above is that it assumes all memories are equally important. This is highly unlikely to be the case; instead, multiple studies show that availability of a memory also reflects its intrinsic importance, even when recency, frequency, and contextual factors are accounted for. For example, stimuli associated with higher monetary rewards are remembered better than those associated with lower monetary rewards [87–89] and emotionally arousing stimuli memorized under stressful conditions tend to be remembered better than stimuli memorized under non-stressful conditions [90]. Additionally, memories of emotional events are thought to persist longer and be more readily available for retrieval than non-emotional memories due to greater activation of the amygdala during memory encoding processes in the hippocampus [7,91].

The role of emotion in memory encoding and retrieval can be accommodated within the classic rational analysis framework, and the prior limitation can be overcome in a simple and coherent way. In the original formulation [4], a memory's availability simply reflected its probability of being re-encoded. If memories differ in terms of importance, a natural metric to allocate resources to memories is the product of a memory's probability and its importance for survival, which can be defined as the emotional *intensity* of the experience at the time of encoding. This product is analogous to the concept of expected utility in decision-making [92],

which is defined as the product of an outcome's probability by its utility. In this case, a memory represents an outcome, its probability is the memory's posterior probability of being re-encoded, and its "utility" is its emotional intensity. Thus defined, the intensity of a memory can be captured by a term, $0 < I(m) < \infty$.

Because rational analysis is framed in odds instead of probabilities, the term *I(m)* will be transformed into an equivalent odds-like formulation and scaled by the mean intensity of all the other memories. Specifically, we will indicate the intensity of a memory *m* as $I(m) / I(\neg m)$, or the ratio between *m*'s intensity and the mean intensity of all other memories that are not *m*, i.e. $I(\neg m) = \sum^{N}_{i \neq m} I(i) / N$. Thus, just like a memory's availability depends on the ratio between a memory's probability and the probabilities of all other memories, its importance depends on the ratio between its intensity and the intensity of all other memories. With these assumptions in place, we can rewrite Eq. (2.1) expressing a memory's activation *A(m)* as its posterior probability multiplied by its intensity $I(m) / I(\neg m)$:

$$
\begin{aligned}
A(m) &= \log\left(\frac{P(m|Q)}{P(\neg m|Q)} \times \frac{I(m)}{I(\neg m)}\right) \\
&= \log\left(\frac{P(m)}{P(\neg m)} \times \frac{P(Q|m)}{P(Q|\neg m)} \times \frac{I(m)}{I(\neg m)}\right) \\
&= \log\left(\frac{P(m)}{P(\neg m)}\right) + \log\left(\frac{P(Q|m)}{P(Q|\neg m)}\right) + \log\left(\frac{I(m)}{I(\neg m)}\right) \\
&= \log\left(\frac{P(m)}{P(\neg m)}\right) + \log\left(\prod_q \frac{P(q|m)}{P(q|\neg m)}\right) + \log\left(\frac{I(m)}{I(\neg m)}\right) \\
&= \log\left(\frac{P(m)}{P(\neg m)}\right) + \sum_q \log\left(\frac{P(q|m)}{P(q|\neg m)}\right) + \log\left(\frac{I(m)}{I(\neg m)}\right) \\
&= B(m) + S(m) + log\, I(m) - log I(\neg m) \qquad (2.6)
\end{aligned}
$$

This Bayesian framework suggests that emotional and traumatic events add a *constant bias* for activation in the form of log *I(m)* and log *I(¬m)*. We predict that this constant bias makes a

memory more likely to be retrieved, even in the absence of contextual cues and in proportion to the perceived intensity of the event during encoding.

2.3 Interpretation of Intensity, *I*

Crucially, the new term $I(m)$ that captures the *intensity* of a memory can be interpreted in reference to emotional processing. A common theme across affective neurosciences is that emotions, at least in their most basic forms, are needed for survival [93–95]. Therefore, events that are associated with different emotions, or to the same emotion but to a different degree, should also differ in their importance to survival. For example, for a prey animal, the memory of a place where a previous encounter with a predator has occurred is more helpful for survival than the memory of the same place if no predator attack had occurred. Thus, the emotional state associated with an episodic memory should translate to its importance.

Although authors disagree on how to categorize emotions, most authors agree that emotions can be placed at least across two dimensions, *valence* and *arousal*. This two dimensional space can be visualized by the c*ircumplex model* of emotions [96]. The term $I(m)$ can be thought of as the norm of the vector in this bidimensional space, capturing an emotion's intensity into a single metric (Figure 2.2). With this interpretation if $I(m)$, valence can be both positive and negative, which fits well into our expanded Bayesian framework because both positive and negative emotions elicit increased neuronal activity during memory encoding and enhance subsequent retrieval of memories [7]. The proposed utility model, therefore, could be used to model the memory dynamics of all emotions.

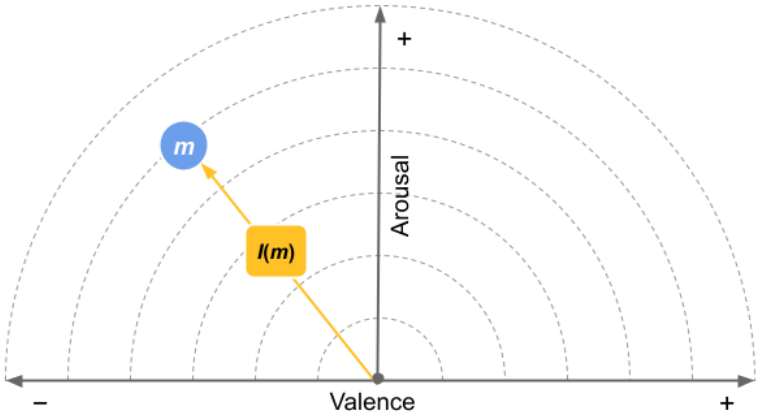

Figure 2.2. An interpretation of the *I(m)* term in Eq. (2.6) within the circumplex model of emotions: *I(m)* represents the combined effects of arousal and valence associated with memory m.

In this project, we are interested in predicting what happens at the extremes of memory and emotion--the place where the magnitude of emotional distress leads to the formation of pathology and psychological disorders. Because fear lies at the most negative valance and highest arousal of the circumplex model of emotions [97] and only negatively valanced emotional experiences produce PTSD pathology, in this thesis memories are referred to as negatively valanced and as 'traumatic', unless otherwise noted. The intensity of a memory *I(m)* therefore represents the degree to which an event was potentially dangerous or traumatic. In addition to framing the analysis of our proposed model in terms of negative valence and PTSD, we will also limit our examination of the neurobiological responses of memory encoding [of the amygdala] to those which are directly connected to PTSD and well understood in neurophysiological terms [98].

The quantity *I(¬m)* represents the average emotional intensity of all memories that are not m. Therefore, if *m* is the traumatic memory, *I(¬m)* is the average emotional intensity of all the other (non-traumatic) memories. In situations in which a large number of memories exist, and

m is the traumatic memory, the term $I(\neg m)$ can be approximated as a constant value of much lower magnitude than $I(m)$. With this approximation, $\log I(m) >> \log I(\neg m)$ and $\log I(\neg m)$ can be ignored.

When retrieving a memory *m* that is not the traumatic memory however, the quantity $I(\neg m)$ is an average that includes the emotional intensity of the traumatic event, and $I(\neg m)$ can no longer be ignored. So, in order to apply the model to task that requires the examination of memories other than the traumatic memory, such as a fear conditioning, we will consider the value $I(\neg m)$ and show that this parameter contributes to the overgeneralization of fear that is symptomatic of PTSD.

2.4 Emotional Re-Appraisal of Events

The simple form of Eq. (2.6) assumes that the emotional intensity $I(m)$ is identical for every single trace of *m*. While the memory of an emotional event can elicit an emotional or physiological response when it is brought back into awareness, realistically each time the memory is recalled, the intensity of 're-experiencing' is of a different, and typically lower, magnitude than the initial event. As noted above and shown in Figure 3.1, the retrieval of a traumatic memory results in the creation of a new trace, and each new trace would have its own intensity *I* that reflects the specific contribution of the amygdala at that moment in time. Thus, a realistic model needs to capture how *I* would change after each subsequent re-encoding of the traumatic memory.

The new *I* associated with the retrieval of a traumatic memory would likely be a combination of the *I* value that has been retrieved and the value that is being experienced as part of the environment. For instance, Juvina et al. [65] have suggested that such adjustments follow the temporal-difference learning algorithm commonly used in Reinforcement Learning.

In this case, we will assume that the emotional intensity of each trace declines with each retrieval, asymptotically reaching the mean intensity value across all memories, $\bar{I}$. A new trace, thus, would have a new intensity value $\gamma[I(m) - \bar{I}]$. The parameter $0 < \gamma < 1$ represents the vividness of a memory's recollection, and $[I(m) - \bar{I}]$ represents the additional intensity of m above the average $\bar{I}$. As the quantity $[I(m) - \bar{I}]$ gets closer and closer to zero, the intensity of a memory approximates the average of all memories.

If *n* traces of *m* are available, the expected value $I(m)$ at the *n*-th retrieval can be estimated as the mean intensity of all traces:

$$I(m) = \sum_{i=1}^{n} \frac{\gamma^i [I(m) - \bar{I}]}{n} \tag{2.7}$$

This quantity can be approximated by the power sum expression:

$$I(m) = \frac{[I(m) - \bar{I}]\ [\gamma\ (1 - \gamma^n)] / (1 - \gamma)}{n} \tag{2.8}$$

Thus defined, the parameter γ accounts for the last remaining factor that affects intrusive memories, that is, vividness of re-experiencing. For low values of γ, the emotional intensity associated with the retrieval of a traumatic memory quickly declines, causing a less vivid re-experience of the event. On the other end, as γ approaches 1, the emotional intensity of a memory declines less and less, and each retrieval is accompanied by a more vivid re-experience of the original event.

2.5 SUMMARY

In this chapter we have introduced the computational framework used to incorporate emotional memories, and specifically traumatic memories, into the rational ACT-R framework. This extension of ACT-R is called Adaptive Control of Thought - Rational Emotional Memory (ACT-REM).

Chapter 3. NEUROBIOLOGICAL INTERPRETATION OF THE MODEL

3.1 NEUROBIOLOGICAL INTERPRETATION OF THE MODEL

The ACT-REM model outlined above can also be interpreted in terms of brain circuitry. In this interpretation, the framework is broadly consistent with the Multiple Trace Theory (MTT) of memory consolidation [32,33], a neural model that posits that every experience forms an episodic trace that is encoded in the hippocampus and that semantic memories are the result of multiple, partially overlapping traces creating stable and accessible patterns of neural activity. The MTT assumes that episodic memories originate from distributed representations that span multiple cortical areas (Figure 3.1).

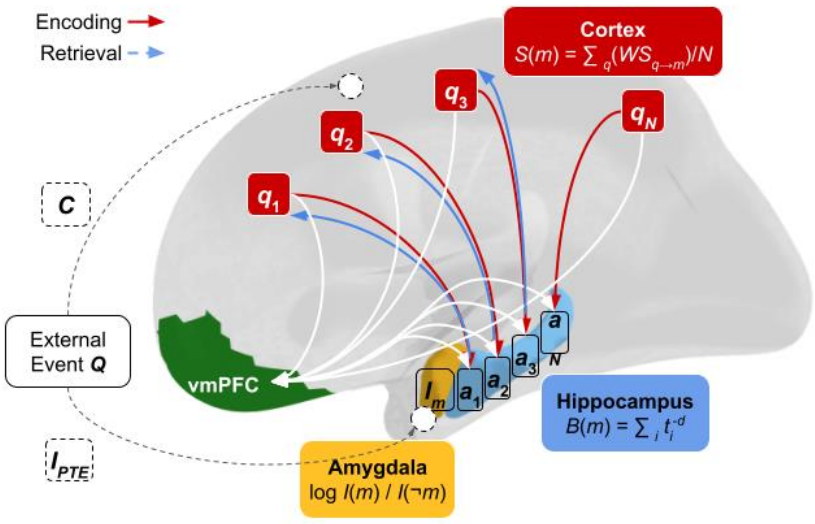

Figure 3.1. Neurobiological interpretation of the model components, their functional contribution, and their main parameters. The red lines represent the encoding process, with a particular context made of cues q_1, q_2, …, q_N being encoded as a new memory m with attributes a_1, a_2, …, a_N. The blue lines represent the retrieval process, with the retrieval of m giving rise to a re-enactment of the original memory and the creation of a new trace for m.

In white, an external event, *Q*, is processed and represents the current context under which a memory will be retrieved. This context is represented by a number of individual sensory cues q_1, $q_2 \ldots q_N$ that are individually represented in different cortical areas. The overlap between the current context and the context under which the traumatic memory was experienced is captured by the parameter *C*. During the encoding phase (red lines), the different cues of the event (q_1, q_2, …, q_N) are encoded in different cortical areas and bound together into a single memory *m* in the hippocampus (as attributes a_1, a_2, …, a_N) through the multiple descending pathways that converge from the cortex through the dentate gyrus.

MTT posits that the hippocampus is the permanent store of episodic memories, and that each encoding episode leaves a permanent trace. During retrieval, the hippocampus reaches a stable pattern that represents a memory trace and, through ascending pathways from the temporal lobe to the cortex, causes the reactivation of the original neurons (blue lines). This reactivation, in turn, might be re-encoded as a second trace. The retrieval of an *appropriate* memory is controlled by the ventromedial prefrontal cortex (green) and is affected by the contributions of three components and their respective neural circuits: (1) the base-level activation *B(m)*, due to the intrinsic processes occurring in the hippocampus (in blue); (2) the spreading activation, due to the signals traveling over the pathways connecting cortical areas and the hippocampus (in red); and (3) an additional emotional boost that conveys the importance of a memory and is due to the activity of the amygdala (in yellow). Base-level activation and spreading activation reflect, therefore, two distinct neural processes. Specifically, base-level activation reflects processes that are internal to the hippocampal network, such as decay or interference due to accumulation of memory traces [36], while spreading activation reflects the mechanism by which cortical inputs might trigger contextual memory retrieval [99].

3.2 Base-level Activation and Hippocampus

In this interpretation, the individual traces that make up every single memory in Eq. (2.6) correspond to the episodic traces of the MTT, and are encoded in the hippocampus (Figure 3.1, blue). The hippocampus receives topologically organized projections from all over the cortex, and (shown via red and white lines in Figure 3.1), and in turn, sends projections back (shown via blue lines in Figure 3.1). The hippocampus is well positioned to perform this encoding because its neurons form a densely 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—a form of content-addressable memory that stores patterns of cortical activity that can then be re-created based on partial inputs. This combination of features makes the hippocampus capable of fast learning and high-storing capacity, making it complementary to the slow learning that takes place in the cortex [100]. The decaying term B_i reflects the progressive weakening of the hippocampal traces, due to either accumulated interference, biological decay [36,101], or a combination of both.

An important biological effect of PTSD is the potential presence of a reduction in hippocampus size. This small to moderate-size effect has been observed numerous times [38–40]. As a memory is retrieved more frequently, its priority increases and its rate of decay decreases. A traumatic memory, however, tends to out-compete more contextually appropriate memories due to the fact it was encoded in a highly emotional state. With each retrieval of the traumatic memory, disproportionately more resources are allocated to it, leading to the further preservation and growth of these unwanted memory intrusions. In this framework, it is proposed that the corresponding changes in hippocampal volume associated with PTSD can be explained

as the natural result of a biological process to efficiently allocate resources to changing memory demands.

In general, it is known that the volume of the hippocampus changes with experience. For instance, in a landmark study [102], cab drivers of London were shown to have larger hippocampal volume than the general population. Additionally, another study showed the volume of the hippocampus co-varies with the years of education [103]. An accepted explanation for this effect is that the volume of the hippocampus reflects the biological investment in storing memories that need to be re-used often [104].

An *efficient* memory storing system would encode neurons such that memories that need to be accessed more frequently use less resources (i.e., fewer cells and/or synapses) than memories that need to be accessed less often [105]. In the Rational Analysis framework described above, memories that are accessed more often have the highest *priors* and, higher *base-level activations*. Knowing the priors of memory utilization, the volume of the hippocampus could then be approximated by a measure of the homogeneity of the distribution of the priors. The degree of efficiency is a function of the information entropy [106] across all M memories in long-term memory. Here, the long-term memory's *information entropy*, H, was utilized:

$$H = -M \times \sum_{i=1}^{M} P(i) \times logP(i) \tag{3.9}$$

This quantity captures how much information is represented in declarative memory once the different probabilities of each memory are taken into account. Consider, for example, the case of two London cab drivers who have memorized the same number of addresses but use them with different probabilities. For one driver, all addresses are equally likely to be retrieved, reflecting the fact that their clients are equally likely to request a ride to all of these locations.

For the second driver, on the other hand, one single address is requested all the time, while all the others are seldom, if ever, requested by clients. Information entropy is high for the first driver because it is impossible to predict which address will be requested by the next client. For the second driver, on the other hand, entropy is low, since one memory is highly predictable and all the other can be ignored. Biologically, the first driver needs to allocate more resources (hippocampal cells) to maintain all these memories than the second, for whom a small number of cells could be used to encode the single memory that predicts most of the clients' rides in their daily routine.

In the proposed neurobiological interpretation of our model (Figure 3.1), the allocation of neural resources to store a memory is divided between the hippocampus and the cortical-hippocampal projections. The allocation of neural resources in the hippocampus reflects a memory's base-level activation $B(m)$ and its intensity $I(m)$. Note the base-level activation $B(m)$ reflects the memory's prior *odds* rather than true probabilities. To translate them into probabilities, base-level activations were normalized across all memories into long-term memory LTM. To calculate the entropy, it is necessary to transform each memory's activation into a corresponding probability value. Because, according to rational analysis, a memory's activation corresponds to its log odds of retrieval, the corresponding probability can be calculated as the ratio of a memory's odds against the odds of all memories:

$$P(m) = \frac{e^{B(m)+I(m)}}{\sum_{j=1}^{M} e^{B(j)+I(j)}} \tag{3.10}$$

Given the theory outlined above, it is hypothesized that increased intrusion occurrence will result in a concurrent decrease in hippocampal size, driven by the altered landscape of memory recall priors, and thereby capturing the relationship between trauma and hippocampal volume.

3.3 Spreading Activation and Cortical Projection

Biologically, the retrieval of memories proceeds in a direction that is symmetric and opposite to the encoding process, with partial inputs from the cortex triggering the re-creation of a pattern of activity in the cortex [32,33,107]. This process and its components are shown in red in Figure 3.1. The spreading activation term S_i captures the role of cortical projection as retrieval cues, with different regions encoding different features and activating different possible traces in the hippocampus. The parameter W determines the strength of top-down control of retrievals. The effect of W can be put into correspondence to the activity of the inferior and medial areas of prefrontal cortex in the control of access to retrieval [108–110]. These processes are colored in white in Figure 3.1.

Biologically, the degree of this control is reflected in the degree of synchronization between medial prefrontal regions and the hippocampus [5,31]. A computational proxy for this measure is the similarity between context Q and the retrieved memory m, with less control associated with lesser similarity.

3.4 Amygdala and Emotional Intensity

The final piece in this circuit is the amygdala, a small nucleus located in front of the hippocampus. As noted above, the amygdala is part of the salience network, a network of regions known to play a role in recognizing possible threatening stimuli and which is hyper-active at rest in PTSD. The amygdala, in particular, plays a fundamental role in processing fear [94], and is bidirectionally connected to the hippocampus. Like the hippocampus, it receives widespread cortical projections from the cortex, which facilitates the immediate recognition of threatening circumstances. In our model, the functional role of amygdala corresponds to the term $I(m)$, which

serves both as an emotional marker for a specific memory *m* and as a cue that boosts its probability of being retrieved.

3.5 Summary and Predictions

In the previous sections, we have outlined a model of how emotional memories can be integrated in an existing computational framework for long-term memory; how an event's emotional intensity boosts the probability of its memory to be retrieved; and how traumatic memories might become intrusive because of this additional boost. Additionally, we have also outlined how this model is compatible with the architecture of the brain circuits supporting emotion, memory, and retrieval, and how this framework could be used to explain hippocampal volume and functional connectivity between prefrontal cortex and the hippocampus.

The rest of this thesis will describe an extensive analysis of this model and of how various factors, manipulated as model parameters, affect the model's outcomes in terms of simulated behavior and simulated brain dynamics.

Chapter 4. METHODS

4.1 Model Implementation

An algorithmic version of the model was implemented in the ACT-R cognitive architecture ACT-R is a natural choice, since its declarative memory system reflects the Bayesian analysis that is the basis of Eq. (2.1) and only minor modifications were needed to implement the new activation equation Eq. (2.6). Furthermore, because ACT-R is currently the most popular cognitive architecture in psychology [62], our code could be easily integrated in other models of memory. Finally, ACT-R is noteworthy because many of its modules and parameters have been

put into correspondence with underlying neural circuits [67,111,112], thus providing an additional level of plausibility to our model. [The entire code for the model, as well as the complete results of the simulations, is available at https://github.com/UWCCDL/PTSD].

4.2 Memory Representations

In ACT-R, memories are represented as "chunks": vector-like structures containing a predefined number of "slots", each of which contains one "attribute". In general, ACT-R gives programmers much latitude to define memories arbitrarily, using the same format for both episodic and semantic memory representations. The focus of this study, however, is explicitly centered on the retrieval of specific *episodic* memories. For this reason, we adapted ACT-R representations in the following ways.

In the model, new memories are arbitrary and randomly generated. The structures of these memories are controlled by two parameters, N and a, which determine the number of slots in a given memory (N, corresponding to the maximum number of contextual cues q) and the number of attributes from which a value for a slot can be chosen from (a, which determines the richness of the internal representations of events).

Although PTEs might occur in familiar environments and conditions (i.e., partner violence), many traumatic events occur in situations, conditions, and environments that are unique and different from the daily life of the agent; for example, for a war veteran, the environment in which the PTE occurred (a war campaign in a foreign country) might have few points of contact with his or her everyday life back home. The degree of exceptionality of the PTE was captured by parametrically varying the pool of attributes that were chosen for the PTE's slots. Specifically, a congruency parameter $0 \leq C \leq 1$ 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 $C = 0$, the PTE is entirely made of unique attributes, while for $C = 1$ the PTE is indistinguishable from the other randomly generated episodes.

4.3 Contextual Effects

In ACT-R, the contextual cues $q_1, q_2, \ldots q_N$ are typically held in separate, content-dependent "buffers", each of which represents a different cortical region and provides its own source of spreading activation. Episodic memory, thus, would be best thought of as an aggregate snapshot of the contents of these buffers. The model was simplified so that events and memories are represented in the same format, as a single vector of N attributes. For the sake of simplicity, contextual information and retrieved memories are allocated to different buffers, identified by the red and blue slots in Figure 3.1 ($q_1, q_2, \ldots q_N$ and $a_1, a_2, \ldots a_N$ respectively). Thus, an event is represented as a context vector Q that is held in the *imaginal* buffer, and the retrieved memory m_T is temporarily placed in the *retrieval* buffer. After the context has been processed and a response has been made, the current Q becomes a new episodic memory that is added to the long-term declarative memory system and might therefore be retrieved in the future.

Since, in our model, all the contextual cues are embedded into a single chunk, ACT-R's spreading activation function was simplified. Specifically, spreading activation flows from each contextual cue q_i to m if the attribute of the i-th slot of m is identical to q.

4.4 Distribution of Contextual and Internally Generated Retrievals

To create a realistic simulation, the model went through several simulated days of activity in which memory retrieval was either guided by external events (which form the context Q to which a response might be made) or by spontaneous recollection. This latter process is an

abstraction for a variety of mental processes that also lead to the retrieval of the PTE, such as rumination, negative appraisal, and even unintentional mind-wandering.

The generation of events that lead to retrievals was modeled as a stochastic process, with the probability of each event being controlled by a gamma distribution over the time of day. Two such distributions were created, one for contextual events and one for the internal unproductive processing events. For a modicum of realism, it was assumed that the contextual events peak in the early morning and are maximally distributed during working hours, while unproductive processing events peak at the end of working hours and extend later in the evenings. Figure 4.1 illustrates the probability densities of the two types of events. Note that, because the cumulative distribution of each curve is 1, the expected number of events of each type during the day is 1. To alter the expected number of events during the day, the probability density curves were multiplied by two scaling factors, E (for the events) and U (for unproductive processing). Thus, the scaling factors can be interpreted as the expected number of events of each type occurring within a day.

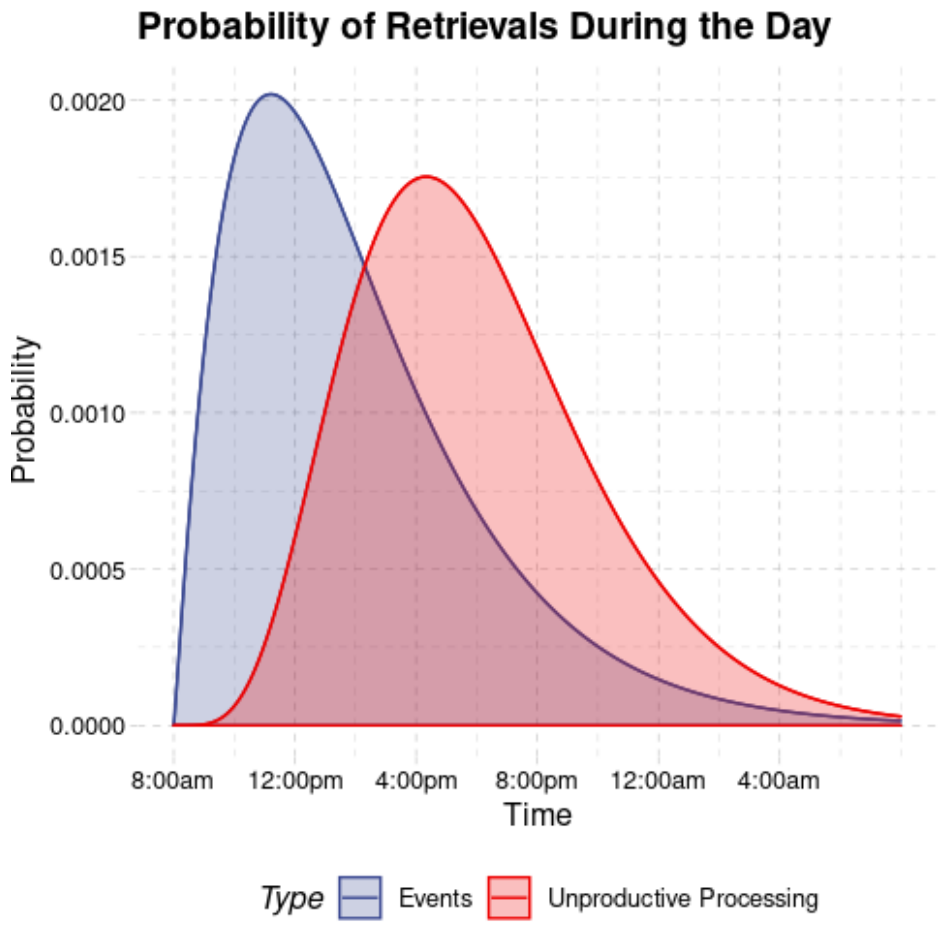

Figure 4.1. Probability density functions of a retrieval happening because of external events (blue) or because of unproductive processing (red) during each simulated day.

4.5 Model Simulations and Data Analysis

Simulations were run by systematically varying the model parameters as shown in Table 4.2. The range of these parameters was chosen as follows. The values of N and a were chosen to create a sufficient variety of different events for the simulations while maintaining the total amount of computations manageable. The values of W were chosen so that its middle value would be equal to N (so that the weight of each cue would be $W/N = 1$ in the spreading activation equation), and the lower and upper values would correspond to 50% decrements or increments of this central value. To capture the effect of context, C the range of was divided into four equal

intervals, and the first four values that divided. The values of *I* were evenly space to include a baseline conditions (*I* = 1) and three conditions (*I* = 20, 40, and 60) in which the net effect of a traumatic event would be comparable but slightly larger than the effect of spreading activation (i.e., log(*I*) = 1.3, 1.6, and 1.8, respectively). The values of *U* were chosen to be either zero, or to match the expected number of retrievals *E*, thus anchoring the range of plausible values. Finally, the values of γ were chosen in analogy to the most common values used, in the modeling literature, to capture the temporal discounting factor of future rewards [113].

Table 4.2. Summary of the simulation parameters

Parameter	**Meaning**	**Value or Range**
N	Number of features in episodic memories (cues in event)	8
a	Number of possible attributes for each feature	4, 6
C	Congruency between environment and PTE	0, 0.25, 0.5, 0.75
W	Cognitive control (Working memory capacity)	4, 8, 12
d	Decay rate	0.5 (standard decay rate)
I_{PTE}	Intensity of PTE	1, 20, 40, 60
E	Frequency of event-driven retrievals	20
U	Frequency of spontaneous retrievals due to unproductive processing	0, 20
γ	Vividness of recalled experience	0.8, 0.9, 0.95

For each combination of parameters in parameter space, the model was run 100 times. Each run spanned 160 simulated days, of which 100 occurred before the PTE to populate the model's long-term memory with a minimal history.

The model was implemented using the ACT-R software, version 7.5 [68]. All of the model simulations were run and maintained through GNU Parallel [114]. The results of the simulations were analyzed with the R language and statistical software [115]. Data were visualized using the

"ggplot2" package [116]; the model's predictions about structural and functional changes in the brain were visualized using the "ggseg" package [117].

4.6 Simulation of Potentially Traumatic Events

To simulate the effects of emotion and trauma on declarative memory, the ACT-R code was augmented with a set of functions that maintain the scalar value $I(m)$ for every memory m created. Every time a new memory is added, its value $I(m)$ is computed and recorded. The value of $I(m)$ is determined on the basis 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 $I(m)$ is set to a predefined value of I_{PTE}, which is parametrically manipulated in the following simulations. Otherwise, the value is drawn from a uniform distribution between 0 and 2, so that the mean value of all the non-traumatic memories is $I(\neg m_{PTE}) = 1$ and, therefore, $\log I(\neg m_{PTE}) = 0$ in Eq. (2.6).

4.7 Model Behavior

The model simulates behavior over a period of time that spans several months, thus stretching the boundaries of a typical ACT-R model's timescale. At midnight of a predefined day, indicated as Day 0, a potentially traumatic event (PTE) is generated and presented to the model. This event is marked as having a greater-than-normal emotional intensity I_{PTE}, whose precise value of I_{PTE} was manipulated during the simulations. The model follows a simple Perceive/Retrieve/Respond loop, in which the model faces external events by retrieving a memory of a previously encountered situation and using this memory to generate a response. The outline of such a procedure is illustrated in Figure 4.2.

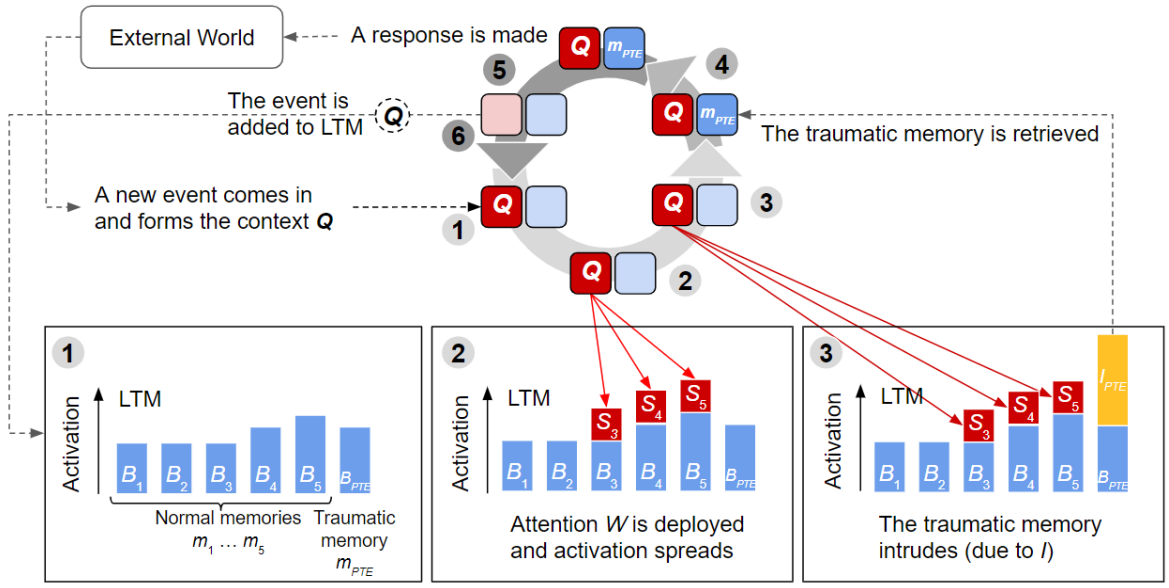

Figure 4.2. Perceive/Retrieve/Respond loop of the model's behavior during a traumatic memory intrusion. The model uses separate buffers to store contextual information (*Q*, red) and the retrieved memory (*m*, blue). (1) Initially, the availability of memories only reflects their base-level activations *B*. After an external event is presented, it is encoded to form the current response context *Q*. (2) Spreading activation from *Q* changes the landscape of the model's long-term memory (LTM), varying the activation level of different memories to reflect their current relevance. (3) The additional emotional intensity I_{PTE} of a traumatic memory m_{PTE} provides an additional activation boost. (4) Because of its additional activation, the traumatic memory m_{PTE} intrudes in the retrieval process even if it is the contextually inappropriate response to the current event. (5) Eventually, the model's memory is used to make a response, and (6) A new memory is formed of the recent event *Q*.

Throughout a series of simulated days, the model is presented with a stream of new events. Each event, in turn, becomes the current context *Q* and its elements become the current contextual cues q_1, $q_2 \dots q_N$. During the *perception* phase (1-3), the model attends to the event, making the cues become the source of spreading activation. The values of *B*(*m*), *S*(*m*), and *I*(*m*) are computed for every element in long-term memory (LTM). *B*(*m*) is easily computed from the

past history of encodings and retrievals of every memory. $S(m)$ is computed by calculating the match between each contextual cue q and the corresponding feature of every other event in LTM. Finally, $I(m)$ is calculated for every event in memory by looking up its corresponding entry in a corresponding table. During the *retrieval* phase (4), the memory with the highest activation value is then selected through a noisy soft-max procedure, and is placed in a special retrieval buffer, marking the end of the retrieval phase. Eventually, during the *response* phase (5-6), a response is performed based on the contents of the retrieval buffer, and the internal representation of the context Q is encoded as a new memory in LTM.

4.8 Dependent Variables

4.8.1 *Behavioral Measures*

During each run of the model, the unintentional retrieval of intrusive memories during a single day was recorded for up to 60 days after the occurrence of the PTE. For each run, two dependent variables were calculated. The first is the mean probability of retrieving intrusive memories over time, from the day in which the PTE occurred to Day 60. This measure represents an average, ideal trajectory for a specific combination of parameters.

The second measure is distribution of recovery trajectories across all runs within a combination of parameters. To calculate this measure, instead of averaging across all of the recovery curves, the recovery curve of each individual instance of the model is classified in one of the four PTSD recovery trajectories defined by Galatzer-Levy et al. 2018 [11], and discussed in the Introduction (Figure 1.1). To categorize each run, the following algorithm was used:

1) First, three values are calculated:

a) The mean probability $P_{baseline}$ of retrieving intrusive memories in the 10 days preceding the PTE, or baseline period. By definition, this is always zero.

b) The mean probability P_{acute} of retrieving intrusive memories in the 10 days following the PTE, or during the acute period.

c) The mean probability $P_{chronic}$ of retrieving intrusive memories in the last ten days of the second month after the PTE (i.e, days 51-60 after PTE), or the chronic period.

2) Then, the model uses these three values to calculates two statistical tests:

a) A *t*-test between the baseline and acute period, or *acute test*.

b) A *t*-test between the acute-period and the chronic period, or *chronic test*.

3) If the *acute test* is significant at $p < .05$ ($P_{acute} > P_{baseline}$), then:

a) If the *chronic test* is also significant at $p < .05$, and $P_{chronic} > P_{acute}$, then we classify the trajectory as *delayed*.

b) If the chronic test is significant at $p < .05$ but $P_{chronic} < P_{acute}$, then we classify the trajectory as *recovery*.

c) If the chronic test is non-significant, then $P_{chronic} \sim P_{acute}$, and we classify the trajectory as *chronic*.

4) If, instead, the *acute test* is *not* significant ($P_{acute} = P_{baseline}$), then:

a) If the *chronic test* is significant at $p < .05$, then $P_{chronic} > P_{acute}$, and we classify the trajectory as *delayed*.

b) If the *chronic test* is also not significant, then $P_{chronic} = P_{baseline} = P_{acute}$, and we classify the trajectory as *resilient*.

4.8.2 *Neurobiological Variables*

A third and final group of dependent variables was collected to see how well the model captures the possible neurobiological effects of traumatic events. These variables were estimated at the chronic period, i.e., days 51-60 after the PTE.

The first neurobiological dependent variable was the predicted change in hippocampal volume. This was calculated using the entropy equation, Eq. (3.9), and averaging the value of H at the end of days 51-60. Specifically, the value of H at the end of each run was compared to the value of H in a baseline condition that would be representative of a healthy, non-traumatized control population. This control condition had parameters that were representative of the general population ($W = 8$, $C = 0.25$, $\gamma = 0.9$; see Results section, below, for how these representative parameters were identified) but was run under no trauma ($I_{PTE} = 1$). The percentage difference of each run's entropy H from this baseline condition was taken as the percentage decrease in hippocampus volume.

The second neurobiological dependent variable was the functional connectivity between the medial prefrontal cortex and the hippocampus. This was calculated by measuring the average similarity between the context Q and the retrieved memory during the chronic period, i.e., days 51-60 after the PTE.

4.9 Modeling Fear Overgeneralization in PTSD

4.9.1 *The Fear Conditioning Paradigm*

To study fear generalization we use the conditional discrimination procedure (abbreviated as *AX*+/*BX*-) developed by Jovanovic et al. [118] that allows for the evaluation of fear potentiation and inhibition of fear in humans. In the *AX*+/*BX*- fear conditioning paradigm,

individuals experience a number of trials in which neutral compound stimuli *AX* and *BX* are either associated with an aversive airblast stimulus (*AX*+) or not (*BX*-). Fear generalization is measured as the probability of a startle response to the presentation of a compound stimulus that was previously not associated with the aversive stimulus (*BX*) and a previously unseen compound stimulus (*AB*). The generalization probability reflects the learned association of *X* or *A* to the aversive stimulus [47,48,118].

This paradigm can easily be accounted for by ACT-REM. We define a startle response as the retrieval of an AX+ memory in the task environment when a cue (either an *AX*, *BX*, or *AB* compound stimuli) is presented. Specifically, when presented with a cue the probability P(startle) of retrieving an *AX*+ memory over a *BX*- memory depends on the relative activations of the two chunks, $A(AX+)$ and $A(BX-)$:

$$P(startle) = \frac{e^{A(AX+)}}{e^{A(AX+)} + e^{A(BX-)}} = \frac{1}{1 + e^{A(BX-) - A(AX+)}} = \frac{1}{1 + e^{-\Delta A}} \quad (4.11)$$

where ΔA is the difference in activation between the aversive and non-aversive stimuli. This difference is due to a variety of factors, including spreading activation *S(m)* and the emotional intensity *I(AX+)* of the aversive stimulus + associated with the compound *AX*. Because *X* and *A* are common to all episodes that contain an aversive stimulus there is a residual probability of retrieving an *AX*+ memory in lieu of a *BX* or *AB* prompt. This is an effect of spreading activation, *S(m)*. The startle reflex that results from the aversive stimulus + produces an elevated emotional intensity for the *AX*+ memory, and thus a nonzero value for log *I(AX+)*. This also increases the probability of retrieving an *AX*+ memory even when not contextually appropriate.

Because no aversive stimulus + is associated with the *BX* compound stimulus, log *I(BX-)* is 0. Figure 4.3 illustrates the components *B(m)*, *S(m)*, and log *I(m)* of Eq. (2.6) and their relationship to the activation of memories for events AX+ and BX- in Eq. (4.11).

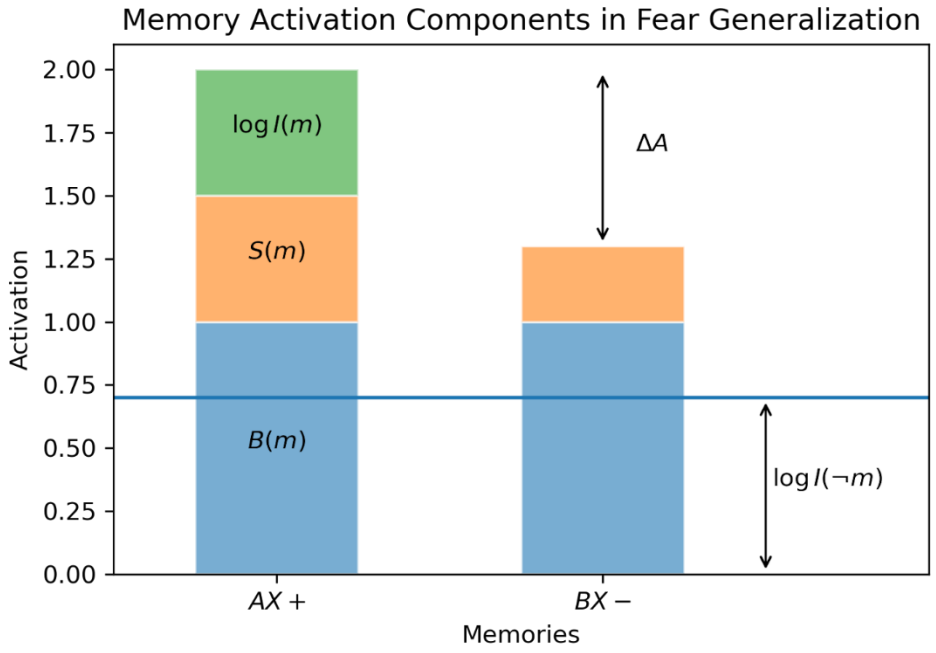

Figure 4.3. Components of memory activation in fear generalization and their relationship to activation of memories for events *AX*+ and *BX*-.

When presenting a neutral *BX* stimulus or novel *AB* stimulus to the model, the spreading activation components are identical between each cue and the *AX*+ startle memory (*A* in *AB*, *X* in *BX*). Therefore, Δ*A* and the probability of retrieving a startle response depends only on the remaining factor log *I(AX+)*, the emotional intensity of the aversive stimulus. Thus, as *I* increases so does its generalization to neutral or novel stimuli, as indicated by the blue line in Figure 4.4.

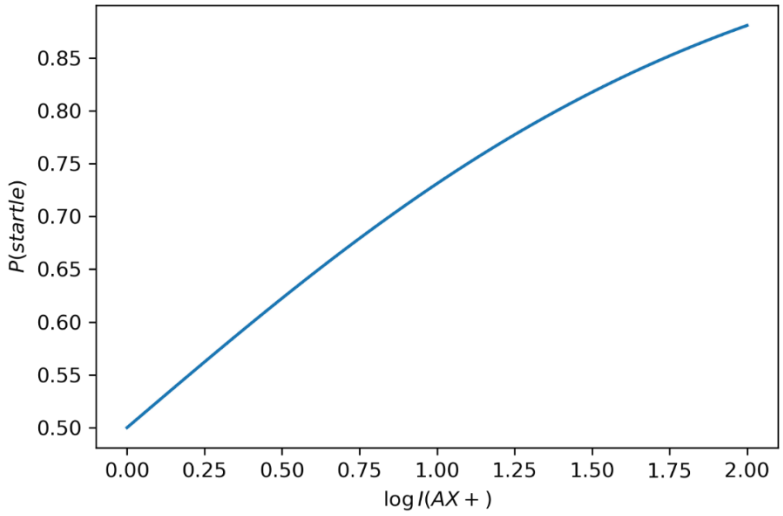

Figure 4.4. The probability of retrieving a startle response P(startle) as a function of log *I (AX+)*.

4.9.2 *Effect of PTSD on Fear Generalization*

As previously noted, an interesting effect of PTSD is that previous trauma enhances future aversive learning [47–49]. This characteristic of PTSD has been shown to increase generalization of a startle response from an aversive learning experience (AX+) to new stimuli (BX or AB) compared to healthy controls with no previous trauma [47,48,118]. In the model framework presented in this thesis, the overgeneralization effect of PTSD can be captured by the term log *I(¬m)*, where *I(¬m)* is the average emotional intensity of all memories other than memory *m*. The value of *I(¬m)* depends on which *m* is retrieved. If *m* is the traumatic memory, *I(¬m)* is the average emotional intensity of all the other (non-traumatic) memories. This value is close to 1 and, therefore, log *I(¬m)* = 0. When we are retrieving a memory *m* that is not the traumatic event, however, the quantity *I(¬m)* is an average that *includes* the emotional intensity of the traumatic event. $I(m_{PTE})$ is typically much larger than that of any other memory and, therefore the log *I(¬m)* term can no longer be ignored (i.e., log *I(¬m)* > 0).

Note in Eq. (2.5), the quantity log *I(¬m)* is subtracted from the activation equation. Therefore, the probability of retrieving the memory of shock (AX+) over non-shock (BX-) episodes changes depending on the value of *I(¬m)*. This change is due to the fact that log *I(¬m)* acts as a positive retrieval threshold. In the case where the activation of BX- and/or AX+ memories fall below the threshold; those memories will *not* be retrieved. Thus, the probability of retrieving a shock event is now the linear combination of the probabilities of the three disjoint events that might happen:

(1) That the activation of both memories AX+ and BX- are above the threshold, and the probability of a startle response depends on the relative activation of the two (See Fig. 4.3, presented again below);

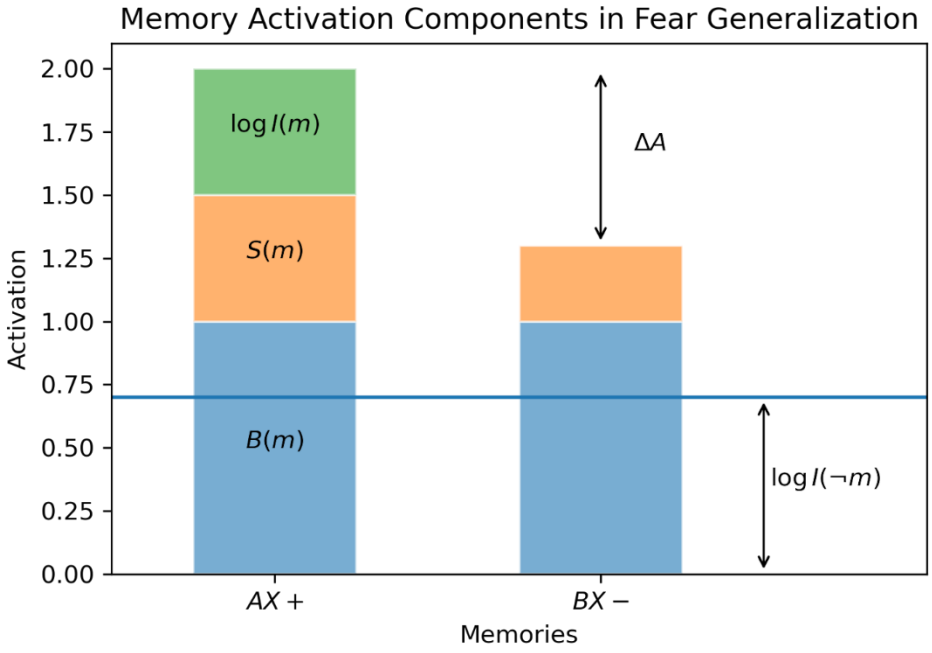

Figure 4.3. Components of memory activation in fear generalization and their relationship to activation of memories for events *AX+* and *BX-*. Threshold log *I(¬m)* is below the activation level for both AX+ and BX- memories.

(2) That AX+ is above the threshold, but BX- is not, so that AX+ is retrieved by default and a startle always happens; and that

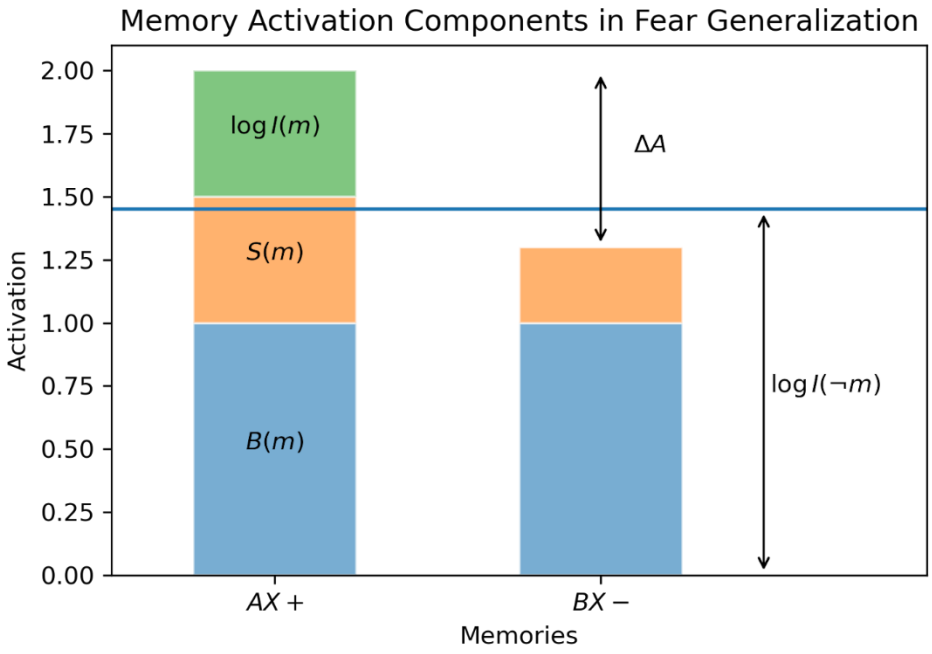

Figure 4.5. Components of memory activation in fear generalization and their relationship to activation of memories for events AX+ and BX-. Threshold log *I(¬m)* is below the activation level for an AX+ memory and above the activation level for a BX- memory.

(3) that AX+ and BX- are below the threshold, under which circumstances no startle happens—at least not due to the retrieval of conditioned stimuli.

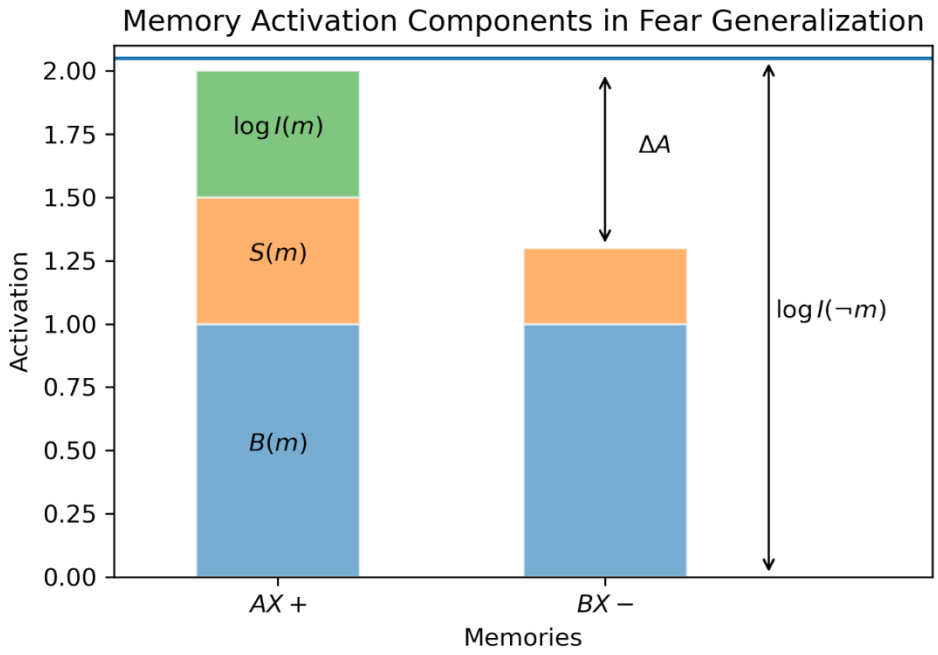

Figure 4.6. Components of memory activation in fear generalization and their relationship to activation of memories for events AX+ and BX-. Threshold log *I(¬m)* is above the activation level for both AX+ and BX- memories.

These three probabilities can be computed as follows. In the first case, the probability of a startle response is simply the probability of retrieving an AX+ stimulus over BX-, and is expressed by Eq. (4.11). The probability of the second case can be computed by thinking of a log *I(¬m)* as a "threshold" *T*, and calculating the probability *P(T)* that the threshold would be higher than the activation of BX-:

$$P(T) = \frac{1}{1 + e^{A(BX-)-T}} \tag{4.12}$$

By analogy, the probability of the third case can be computed as:

$$P(T) = \frac{1}{1 + e^{A(AX+)-T}} \tag{4.13}$$

Because these three events are disjointed, the total probability can be calculated as:

$$P(startle) = \frac{1}{1 + e^{-\Delta A}} + \frac{1}{1 + e^{A(BX-)-T}} - \frac{1}{1 + e^{A(AX+)-T}} \tag{4.14}$$

4.9.3 *Methods*

The original ACT-REM code was modified to replicate a standard version of the AX+, BX- paradigm. In this version, eight "A" stimuli, identified as A1, A2 … A8, were associated with the same abstract stimulus "X" and a shock stimulus "+". Each presentation of the stimulus was encoded as a memory trace (a "chunk" in ACT-R) with three components: the A-type stimulus, the X-type stimulus, and the presence (+) or absence (-) of a shock. As in ACT-REM, each shock stimulus was associated with an emotional intensity value *I*, which was set to 1 (so that log $I = 0$) for the non-shock stimulus. For each simulation, a single test stimulus was presented, consisting of either an AX cue (one of the previously experienced compound stimuli associated with a shock), a BX cue (a compound stimulus previously not associated with a shock), and an AB stimulus (a new component stimulus made of one A-type and one B-type stimulus). Additionally, a 'startle probe' parameter P was added to both AX and BX cues to replicate the presence of the 100-dB startle probe present in the conditioning phase of *AX+/BX-* task implemented by Jovanovic et al. [48]. A 'noise alone' condition, NA, presented in the analysis measures the potentiated startle response to the startle probe, P.

To test the exactness of our mathematical analysis, we conducted a numerical Monte Carlo simulation using the code made available at https://github.com/UWCCDL/PTSD.

Chapter 5. RESULTS

5.1 Memory Intrusions and Recovery Trajectories

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 the number of intrusive memory retrievals per simulated day, with the entire simulation (T_{max}) extending to approximately two months after the occurrence of the PTE. Figure 5.1 provides an overview of the results across all parameter space. First, as expected the model does indeed show worse clinical outcomes in response to more traumatic events. Figure 5.1, below, shows the daily incidence of traumatic memories. A 3x60 analysis of variance (ANOVA), using emotional intensity I and the days after PTE as factors, revealed that I had a significant effect on the relative frequency of experiencing traumatic memories in the days following a traumatic event [$F(2, 1295345) = 37{,}115.6$, $p < .0001$], with higher values of I corresponding to higher relative frequency of memory intrusions. Furthermore, I interacted significantly with the day [$F(118, 1295345) = 10.3$, $p < .0001$], resulting in different recovery trajectories.

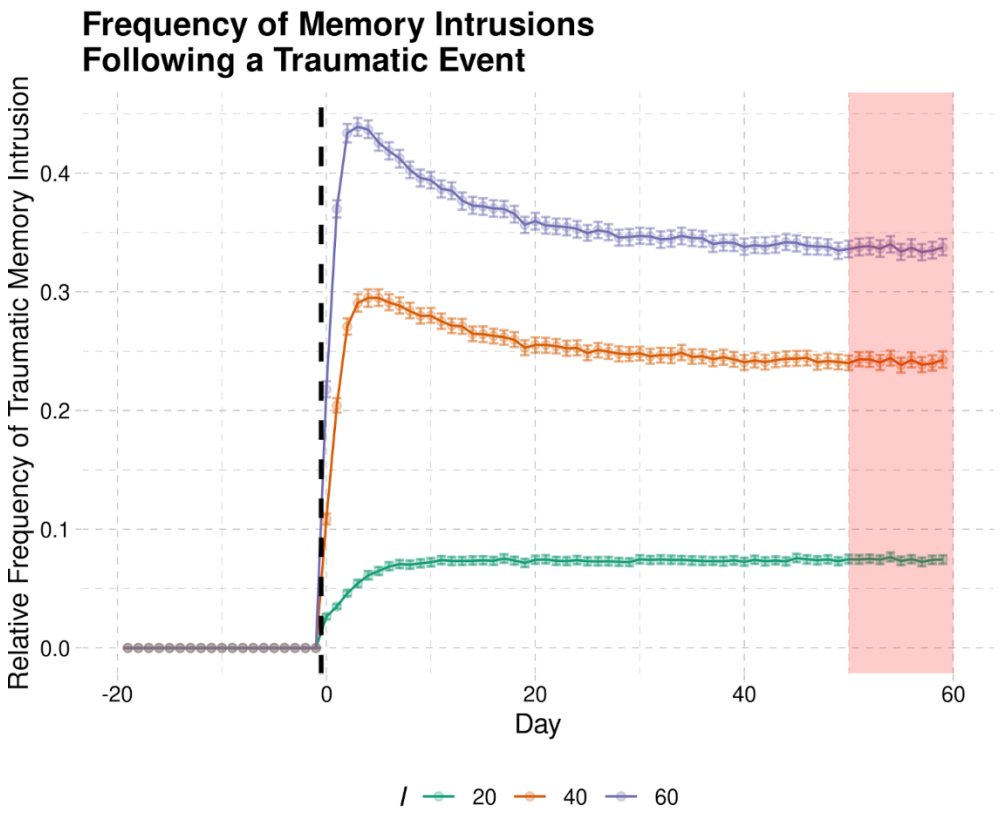

Figure 5.1. Predicted increase in memory intrusion following a PTE on Day 0 (black dashed line) as a function of emotional intensity *I*. Error bars represent standard errors of the mean; the shaded red area marks the time interval in which the hippocampal volume was calculated.

Figure 5.2 shows the predicted number of memory intrusion across all simulated parameters, divided by the type of recovery trajectories of the corresponding model runs. Neither the proportion of trajectories nor the specific number of traumatic memory intrusions in Figure 5.2 are representative of the observed effects in the general population. This is because the parameter values were not selected in a way that was empirically informed---in fact, part of the goal of this study is to identify empirical parameter values and ranges that better characterize the general population.

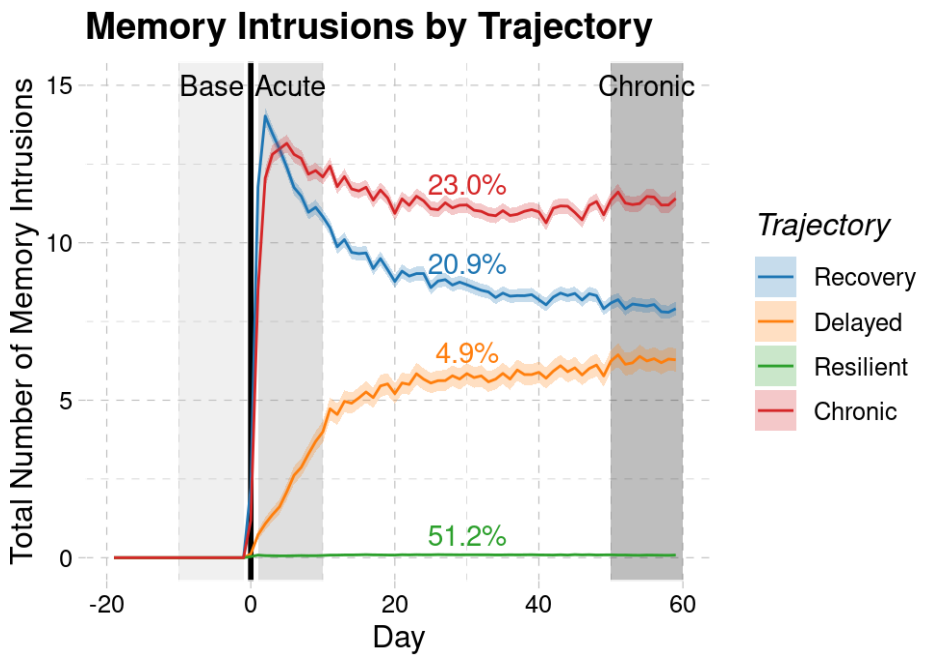

Figure 5.2. Average timecourse and of traumatic memory intrusions and relative trajectory prevalence for each trajectory type, calculated across all of the model parameter values manipulated in the simulations. The shaded areas represent the time intervals at which the frequency of traumatic memory intrusions is measured to identify each model's run trajectory.

To this aim, we examined every combination of parameters to identify the specific set whose proportion of trajectories best represents what is observed in the literature. These proportions were obtained from the meta-analysis reported by Galatzer-Ivy et al. [11], which gives an estimated prevalence for the Resilient (65.7%), Recovery (20.8%), Chronic (10.6%) and Delayed trajectories (8.9%) across 54 studies and 76,435 participants. It was found that, for a particular combination of model parameters (C = 0.25, gamma = 0.9, I_{PTE} = 40, W = 8, and U = 0), the predicted trajectories closely matched the percentages reported in the meta-analysis. The mean trajectories and their relative percentages are depicted in Figure 5.3. The percentage of trajectories is not statistically different from what emerged in the meta-analysis ($\chi^2(3) = 7.46$, $p > .05$) and almost identical to those observed in some recent studies [119].

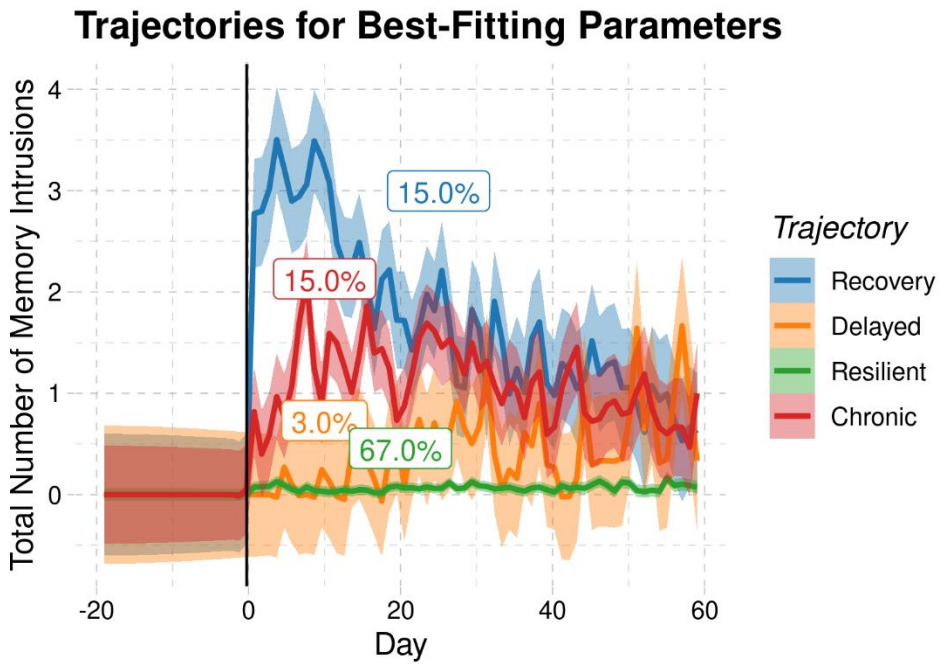

Figure 5.3. Percentages and timecourses of the different trajectories for the combination of parameters that best matches the prevalence of trajectories in Galatzer-Ivy et al. [11].

Having identified a representative set of parameters, we can now examine how systematically varying parameter values that represent different risk factors affect the timecourse and severity of memory intrusions in PTSD, and whether these changes are consistent with the direction of the predicted effects in Table 1.1 and the empirical findings reported in the literature.

5.2 Effects of Environmental Factors

The environmental factors in our simulations are the intensity I_{PTE} of the initial traumatic event and the contextual similarity C between the PTE and the current context Q under which a retrieval process is initiated. To examine these effects, an analysis was conducted on all simulation runs in which the values of idiographic parameters W and γ and the unproductive

processing parameter *U* were kept at the population-level representative values of Figure 5.2, while *I* and *C* were varied at all possible values, presented in Table 4.2.

Figure 5.4 illustrates the mean timecourses of traumatic memory intrusions across different values of I_{PTE} and *C*.

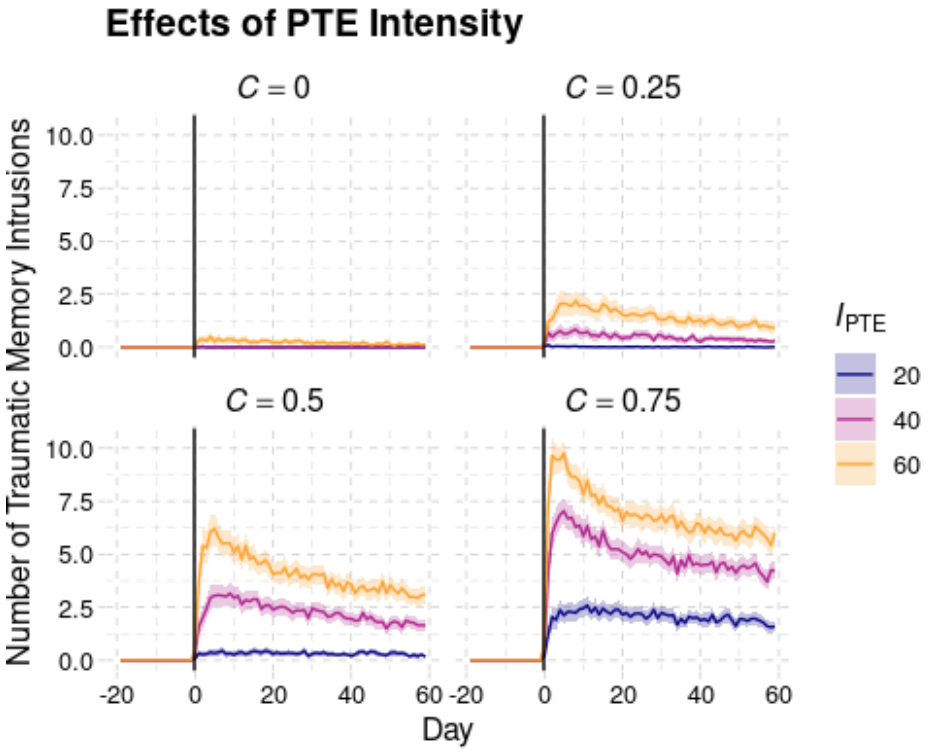

Figure 5.4. Predicted daily number of intrusions following a traumatic event of different intensity (I_{PTE}) under varying values of contextual similarity (C).

A number of effects are consistent with the classic results in the literature. In general the frequency of memory intrusions is higher immediately after experiencing the traumatic event, and tends to taper off over time; this fact is consistent with the prevalence of spontaneous recovery across studies [11]. The frequency of memory intrusions is systematically larger after events of greater traumatic intensity. Additionally, and unsurprisingly, the number of intrusions

tends to be higher when the environment contains a greater number of multiple cues associated with the traumatic event, as modeled by the contextual similarity parameter *C*.

To systematically investigate the effects of *I* and *C*, the average number of memory intrusions *N* at the end of the simulated period of time (Days 51-60) was modeled through a generalized fixed-effect linear model of the form $N \sim \beta_0 + \beta_I I_{PTE} + \beta_C C + \beta_{I,C} I_{PTE} C + \varepsilon$. Because the number of intrusions *N* is a count variable, the statistical model used a Poisson rather than Gaussian link function. The statistical analysis shows a significant effect of both the PTE intensity (standardized $\beta_I = 0.005$, $p < 0.001$) and context (standardized $\beta_C = 0.008$, $p < 0.001$). Their interaction was also significant with a negative parameter value ($\beta_{I,C} = -0.003$, $p < 0.001$).

Although they provide important quantitative information, each of the timecourses of Figure 5.2 is averaged from hundreds of individual runs of the model, each of which would exhibit its own recovery trajectory. We predicted that both the intensity of the traumatic event and the proportion of traumatic cues in the environment would have significant effects on the model's recovery trajectories, with the number of resilient trajectories declining as the values of *I* and *C* increase. The relative distribution of the other three types of trajectories (Recovery, Chronic, and Delayed), however, is harder to predict because of the possible interactions between factors and the fact that each trajectory's prevalence is constrained by the relative frequency of the others. Indeed, this is precisely the type of situation in which computer simulations might provide insight into the balance of forces that affect recovery. Figure 5.5 depicts the prevalence of each of the four trajectories for each combination of I_{PTE} and *C*.

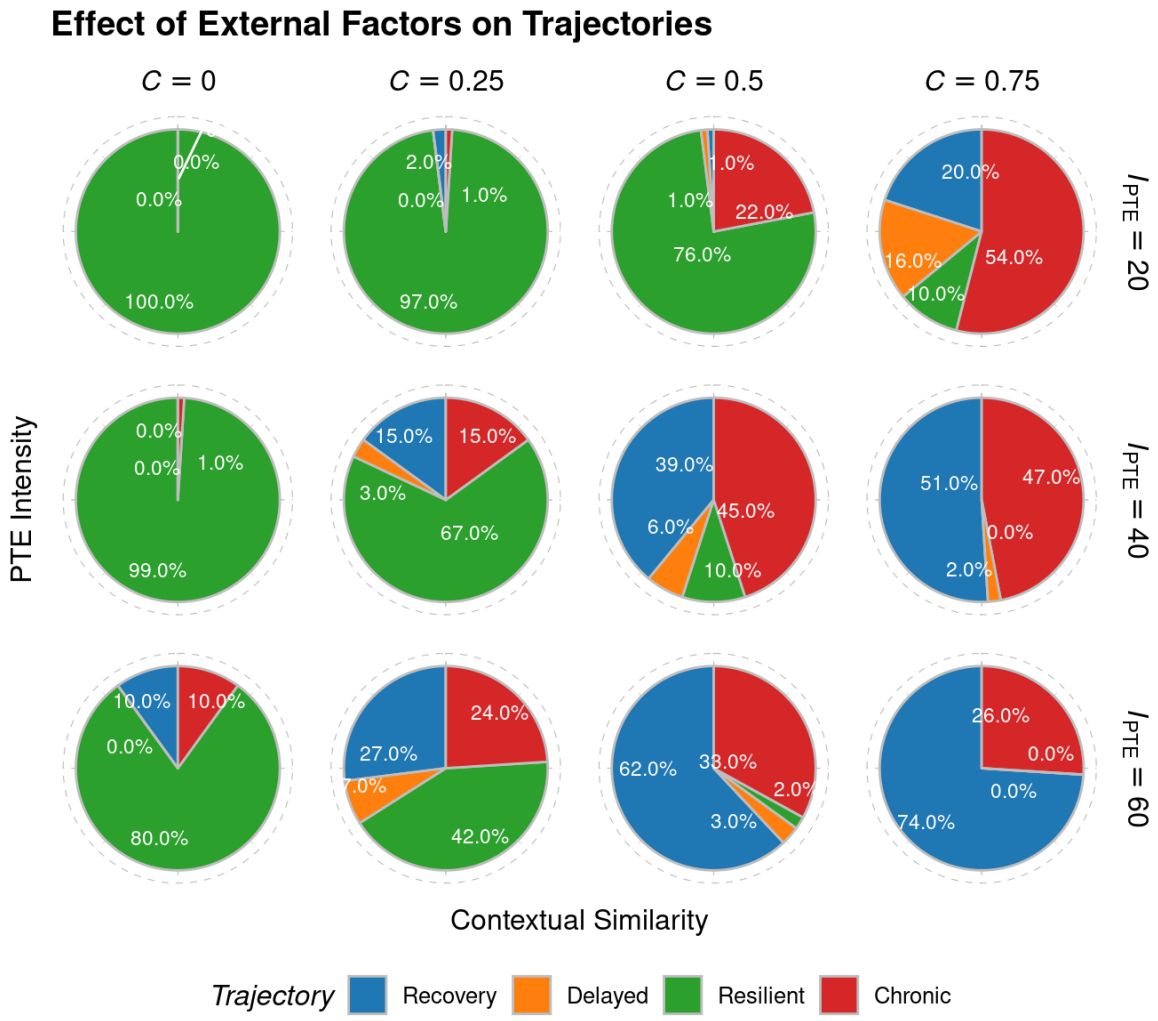

Figure 5.5. Relative proportion of the four fundamental recovery trajectories (Figure 5.2) under different values of the intensity of the traumatic event (PTE) and the contextual similarity between the traumatic event and daily life.

The relative distribution of the four trajectories was analyzed using a MANOVA model. Note that, because the prevalence of the trajectories is bound to sum up to 100%, this statistical model has only three effective dependent variables, since the fourth trajectory's prevalence is completely determined once the first three are known. For the purpose of our analysis, the three dependent variables were the prevalence of the Resilient, Recovery, and Delayed trajectories; the statistical results would not change if any other set of three trajectories were considered. The

MANOVA analysis revealed that that both the intensity of the PTE (Pillai's trace = 0.25, $F(3, 1196) = 133.05$, $p < 0.0001$) and its contextual similarity C (Pillai's trace = 0.54, $F(3, 1196) = 460.64$, $p < 0.0001$) as well as their interaction (Pillai's trace = 0.06, $F(3, 1196) = 24.93$, $p < 0.0001$) significantly affected the distribution of the trajectories.

Because an increase in the prevalence of one trajectory must occur at the expense of the others, as can be seen in Figure 5.5, the relative proportion of each trajectory is different for the different environmental factors. For instance, the Resilient trajectory steadily declines as the values of I_{PTE} and C increase, while the Delayed trajectory waxes and wanes.

To further analyze this effect, four separate analyses were conducted, each of which estimated the probability that a run of the model under different conditions exhibited a specific trajectory. All of the analyses were implemented as a generalized linear model of the form $T = \beta_0 + \beta_I I_{PTE} + \beta_C C + \beta_{I,C} I_{PTE} C + \varepsilon$, with T being a binary variable (coded as $T = 1$ if the specific run exhibits the target trajectory and $T = 0$ otherwise) and the model using a binomial link function.

In the case of the Resilient trajectory (green in Figure 5.5), both the intensity of the PTE (standardized $\beta_I = -3.41$, $p < 0.00001$) and its contextual similarity ($\beta_C = -5.78$, $p < 0.00001$) had significant and negative effects. Furthermore, their interaction did not reach significance ($\beta_{I,C} = -0.39$, $p > 0.59$), implying that their effects were additive. The other three trajectories, however, show a more complicated pattern. The main effects of both factors were significant and positive for all three trajectories (standardized $\beta_I > 2.94$, $p < 0.001$; $\beta_C > 3.28$, $p < 0.0001$); this reflects the deeper and longer-lasting effects of the PTE, which increases the probability of Recovery, Delayed, and Chronic trajectories while decreasing the share of Resilient ones. The interaction term, however, was significant only for the Delayed and Chronic trajectories. In both cases, the

interaction has a negative sign ($\beta_{I,C} < -4.27$, $p < 0.0001$) implying that, as the intensity of the PTE and its contextual similarity increases, the probability of these trajectories might actually decrease (as it does in Figure 5.5). It is also worth noting that, when the environmental context shares very few attributes with the original traumatic event (i.e., $C = 0.0$ or 0.25), resilient trajectories remain common even for high levels of traumatic intensity. This is compatible with the potentially paradoxical finding that resilient trajectories are more common in combat-traumatized veterans, despite the severity of the events witnessed [11].

In summary, the model's simulations largely reflect the patterns observed in the traumatic stress literature, showing consistent effects of both the intensity of the trauma and the presence of traumatic cues in the environment as powerful drivers of intrusive memories and pathological trajectories.

5.3 Effects of Idiographic Factors

Next, we will examine the effects of individual idiographic factors, which in our model are abstracted in the cognitive control parameter W and the intensity of relived experience γ. As shown in Figure 5.6, both factors significantly affected the onset and timecourse of intrusive memories.

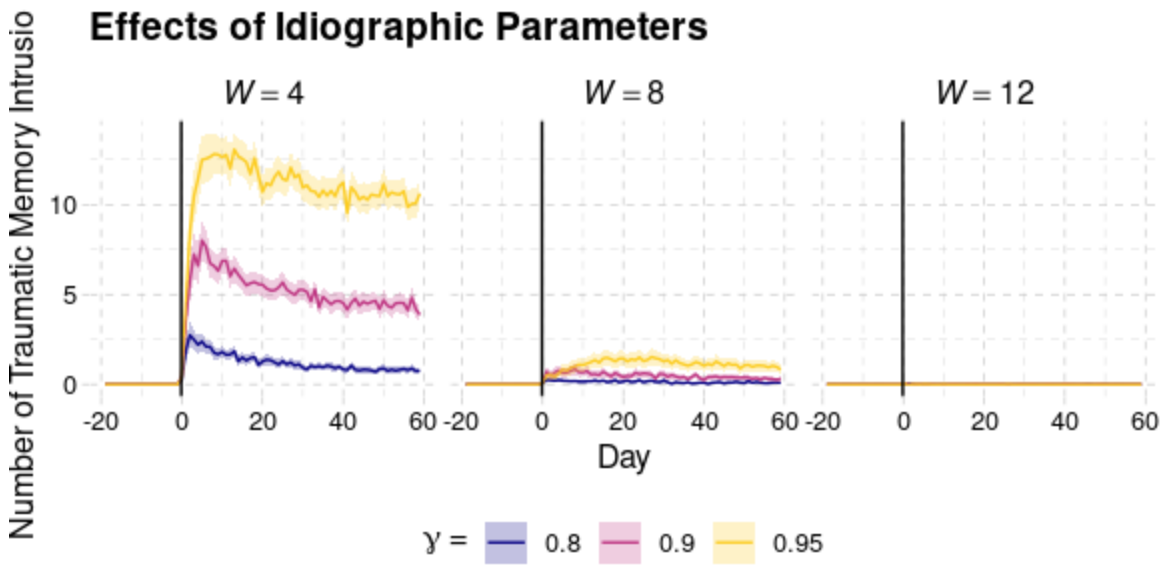

Figure 5.6. Effects of idiographic factors of cognitive control (*W*) and intrusive memory intensity (γ) on daily memory intrusions.

In general, the effects of the cognitive control factor *W* were much stronger than expected, with the model registering no intrusive memories when *W* = 12. This suggests that the range chosen for *W* was unrealistically large, a fact that will need to be taken into consideration in future simulation studies.

As in the case of external factors, the effects of the *W* and γ on the number of traumatic events was analyzed by through a generalized linear model of the form $N = \beta_0 + \beta_W W + \beta_\gamma \gamma + \beta_{W,\gamma} W\gamma + \varepsilon$ with a Poisson link function, with *N* being the cumulative sum of intrusive memories in Days 51-60 after the PTE. As expected, the analysis showed that increased levels of cognitive control were associated with less frequent intrusive memories (β_W = -0.005, p = 0.01), while increased levels of vividness were associated with more frequent intrusions (β_γ = 0.002, p < 0.0001). The two factors did not interact with each other ($\beta_{W,\gamma}$ = -0.0002, p > 0.91).

As before, the prevalence of trajectories was analyzed with a MANOVA model, using the proportion of Resilient, Recovery, and Delayed trajectories as the dependent variables (the

proportion of Chronic trajectories was determined once the other three are known) and the values of W and γ as the independent variables. Both W (Pillai's trace = 0.62, $F(3, 896) = 476.34$, $p < 0.0001$) and γ (Pillai's trace = 0.08, $F(3, 896) = 26.56$, $p < 0.0001$), as well as their interaction (Pillai's trace = 0.05, $F(3, 896) = 14.97$, $p < 0.0001$), were statistically significant, with W explaining the most variance, as indicated by its large Pillai's trace value.

Four separate analyses were conducted on the probability of each model run resulting in one of the four trajectories; all of them used a generalized linear model with a logistic link function. Because of the large variance in the proportion of trajectories (due to the massive effects of W: Figure 5.7), none of the interaction terms were significant. Thus, we analyzed all four trajectories with a simpler generalized linear model of the form $N = \beta_0 + \beta_W W + \beta_\gamma \gamma + \varepsilon$. Increases in the cognitive control parameter W were found to be significantly associated with increases in the Resilient trajectory (standardized $\beta_W = 6.86$, $p < 0.00001$) and with decreases in the probability of Recovery, Delayed, and Chronic trajectories (standardized $\beta_W < -3.63$, $p < 0.00001$). This is consistent with the hypothesis that executive control increases resilience to traumatic events, likely because of the increased ability to control emotional memories [120]. Increases in the recollection vividness γ, on the other hand, were found to be significantly associated with increases in the occurrences of Chronic and Delayed trajectories (standardized $\beta_\gamma > 1.43$, $p < 0.00001$) and decreases in the occurrences of Recovery and Resilient trajectories (standardized $\beta\gamma < -0.90$, $p < 0.001$). This is consistent with our hypothesis that recollection vividness negatively affects recovery by perpetuating the lingering effects of traumatic memories.

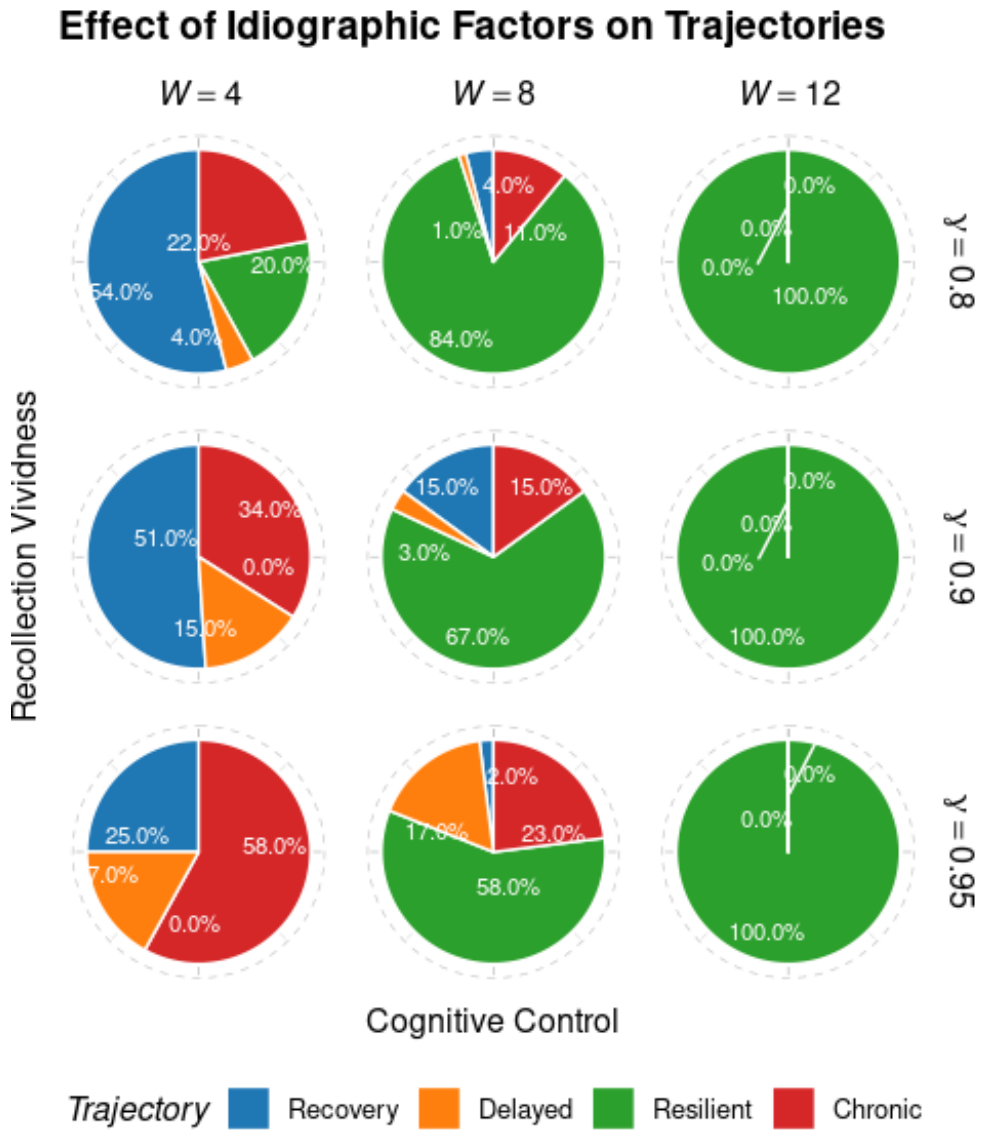

Figure 5.7. Effects of idiographic factors on the prevalence of recovery trajectories.

In summary, both lower cognitive control *W* and higher intensity of intrusive memories γ were associated with more intrusive memories and generally more pathological outcome trajectories. The large effect of *W* is likely due to the fact that the range of parameter values examined was too large. However, while the distribution of trajectories for very high values of *W*

seem unrealistic, the distribution of trajectories and the severity of memory intrusions for very low values of *W* is compatible with those found in cases of intellectual disability [121].

5.4 Effects of Unproductive Processing

Finally, we examined the effects of the presence of unproductive processing, as captured by the parameter *U*. A generalized single-factor linear model with a Poisson link function found that *U* greatly increased the number of daily intrusive memories (Figure 5.8; standardized β_R = 0.003, $p < 0.00001$).

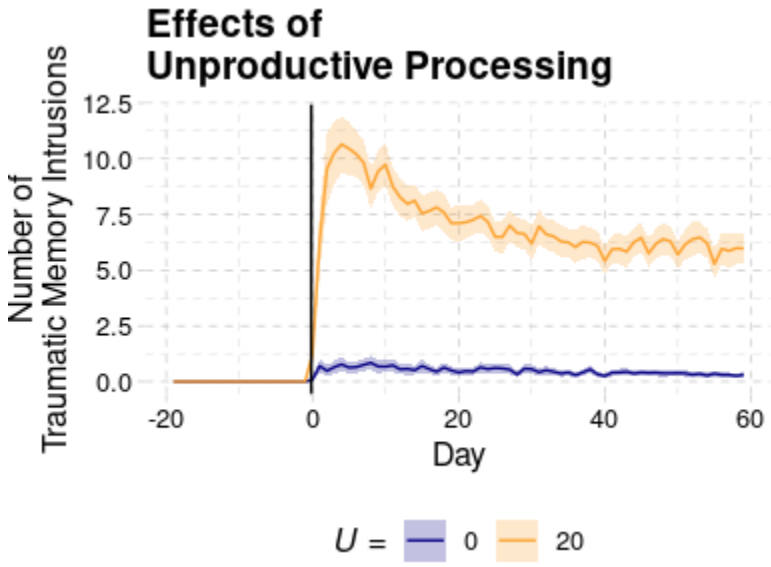

Figure 5.8. Effects of unproductive processing on the frequency of daily memory intrusions.

A MANOVA found that it also significantly affected the distribution of trajectories (Figure 5.9: Pillai's trace = 0.45; $F(1, 196) = 55.09$, $p < 0.0001$).

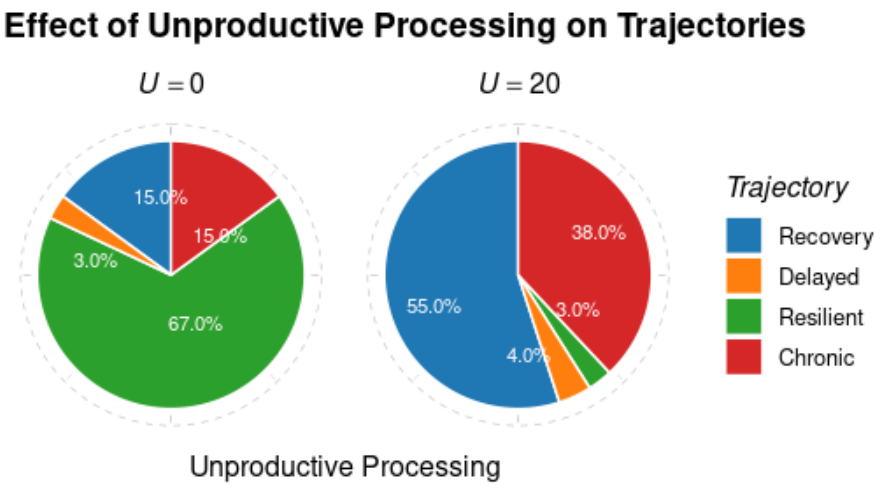

Figure 5.9. Effects of unproductive processing on the proportion of recovery trajectories.

Separate analysis of for each trajectory, using a generalized linear model with a logistic link function, showed that unproductive processing negatively affected the probability of Resilient trajectory (standardized β_R = -4.38, $p < 0.00001$), significantly increased the probability of occurrence of Chronic and Recovery trajectories (standardized $\beta_R > 1.41$, $p < 0.0005$), but had no effect on the Delayed trajectory (standardized $\beta_R = 0.81$, $p > 0.70$).

5.5 Neurobiological Effects of PTSD

In addition to the behavioral effects, the model also generated predictions about how intrusive memories affect the brain. Specifically, we will focus on PTSD-related changes in the size of the hippocampus and in its functional connectivity to the prefrontal cortex. These two measures capture consistent and key abnormalities reported in the literature, and can be approximated as the value of Shannon's information entropy H across all memories, Eq. (3.9), and the mean value of the similarity between the retrieved memory and contextual cues, respectively.

5.5.1 *Hippocampus Volume*

Having established that our model succeeds in capturing individual differences in recovery trajectories for PTSD, the results were further examined to estimate the effects of traumatic stress on hippocampal volume. It was observed, across all parameters, that there was general reduction of simulated hippocampal volume, ranging from zero to 33.89% with a mean decrease of 7.35% [$t(21599) = 140.83$, $p < .0001$]. Both the mean decrease and the range of variation match the results of existing meta-analyses. For example, in Smith's 2005 review of structural MRI studies [122], the decrease in hippocampal volume ranged between 0 and 44% with a mean of 6.9%. A second question was whether the severity of the reduction was predicted by the severity of trauma. To this end, the model simulation results suggest that the emotional intensity I was a significant predictor of hippocampal volume reduction [$F(2, 21594) = 774.7$, $p < .0001$], with the decrease in hippocampal volume growing with greater values of I (all pairwise comparisons significant at $p < .0001$, Bonferroni corrected). This is shown in Figure 5.10, which shows the distributions of predicted decreases of hippocampal volumes in the simulations, visualized (as violin plots) separately for different values of intensity I.

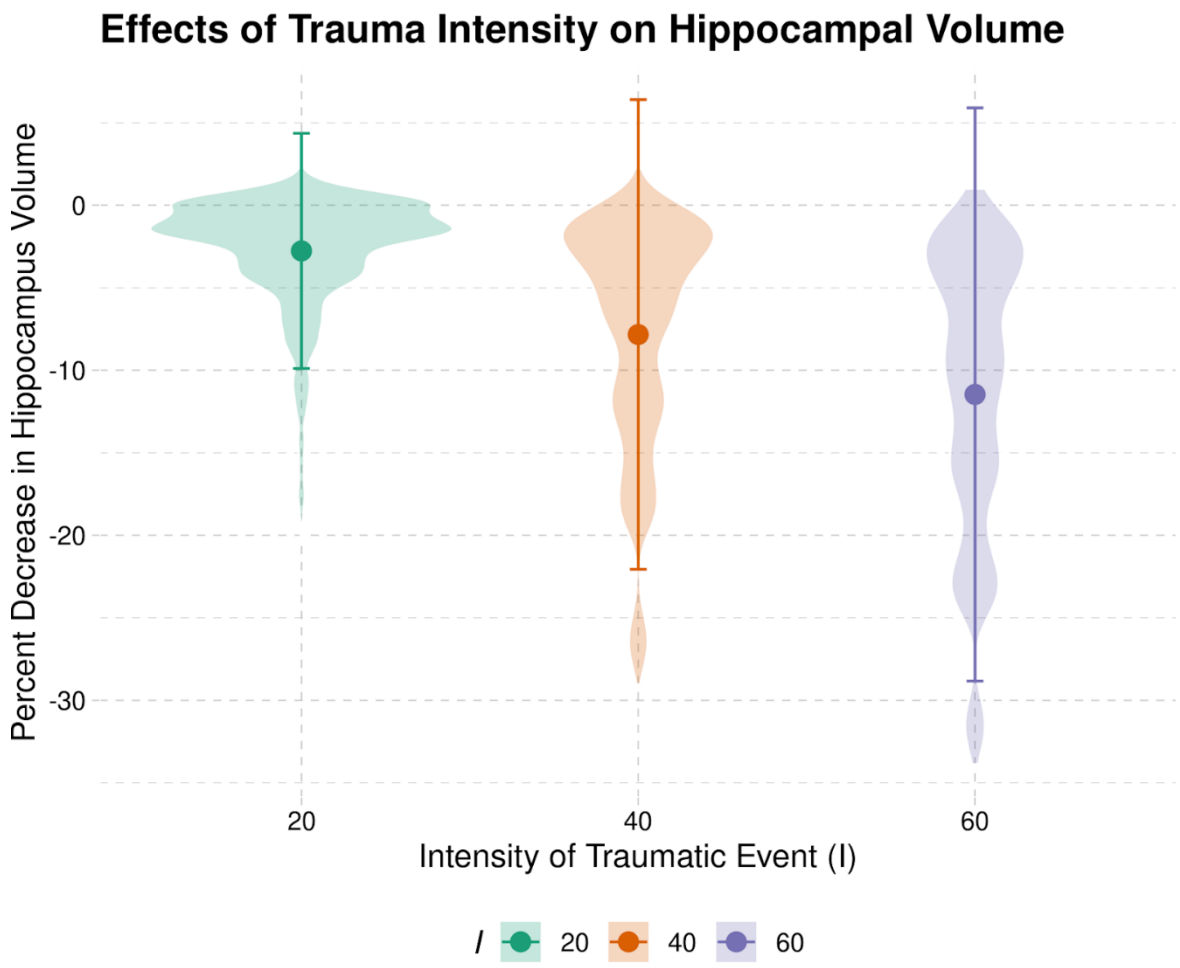

Figure 5.10. Effect of trauma intensity *I* on hippocampal volume. The violin plots represent the distribution densities of model runs resulting in the corresponding decreases of hippocampal volumes. Solid circles and lines represent means +/- standard deviation. In the control condition, the hippocampal decrease is zero.

Although the clinical literature typically reports the prevalence of PTSD symptoms and recovery trajectories following a particular type of traumatic event, neuroimaging studies of PTSD largely focus on differences *between* groups of individuals who already exhibit different symptoms. They could be, for example, individuals suffering from PTSD vs. healthy matched controls with no history of trauma or, more commonly, individuals exhibiting PTSD symptoms vs individuals *not* exhibiting PTSD symptoms after a similar, traumatic event. Because in these

types of studies, the two groups are matched with respect to some of the factors of interest (such as the intensity of the traumatic event or individual differences in executive function, corresponding to our parameters *I* and *W*), it would be uninformative to conduct an analysis of the effects of these factors.

It is also problematic to consider the traumatic events that two different groups have been exposed to as "similar". As noted above, the intensity of a traumatic event interacts with the characteristics of the individual and group assignments often fail to account for other important confounding factors, such as individual trauma history. Furthermore, in the majority of studies, no information is given about the specific recovery trajectory of individuals in the experimental groups. This is problematic because, while they may currently exhibit PTSD symptoms their recovery trajectories could be very different, they could be on a recovery course, chronically stable, or they could have had a delayed onset of symptoms (Figure 5.2).

To conduct a proper analysis that is more comparable with the literature, we analyzed our model's neurobiological dependent variables using each run's *end-point trajectory*, that is, whether each model run's intrusive memories on days 51-60 after the PTE is significantly greater than zero or not in a *t*-test. This lumps models runs that either never developed intrusive memories (Resilient) or successfully recovered after developing them (Recovery) into the "Control" group and model runs that either experienced intrusive memories throughout the run (Chronic) or developed them later on (Delayed) into the "PTSD" group.

Figure 5.11 illustrates the results of this analysis for hippocampal volume. The simulated PTSD group's change in hippocampus volume (M = -8.63%, SD = 7.82%) was significantly larger than the traumatized control group's (M = -0.04%, SD = 1.47%; Welch two-sample test

t(12754) = 92.42, *p* < 0.0001), with the control group's change in hippocampus being not significantly different than zero (one-sample *t*(3097) = -1.50, *p* > 0.13).

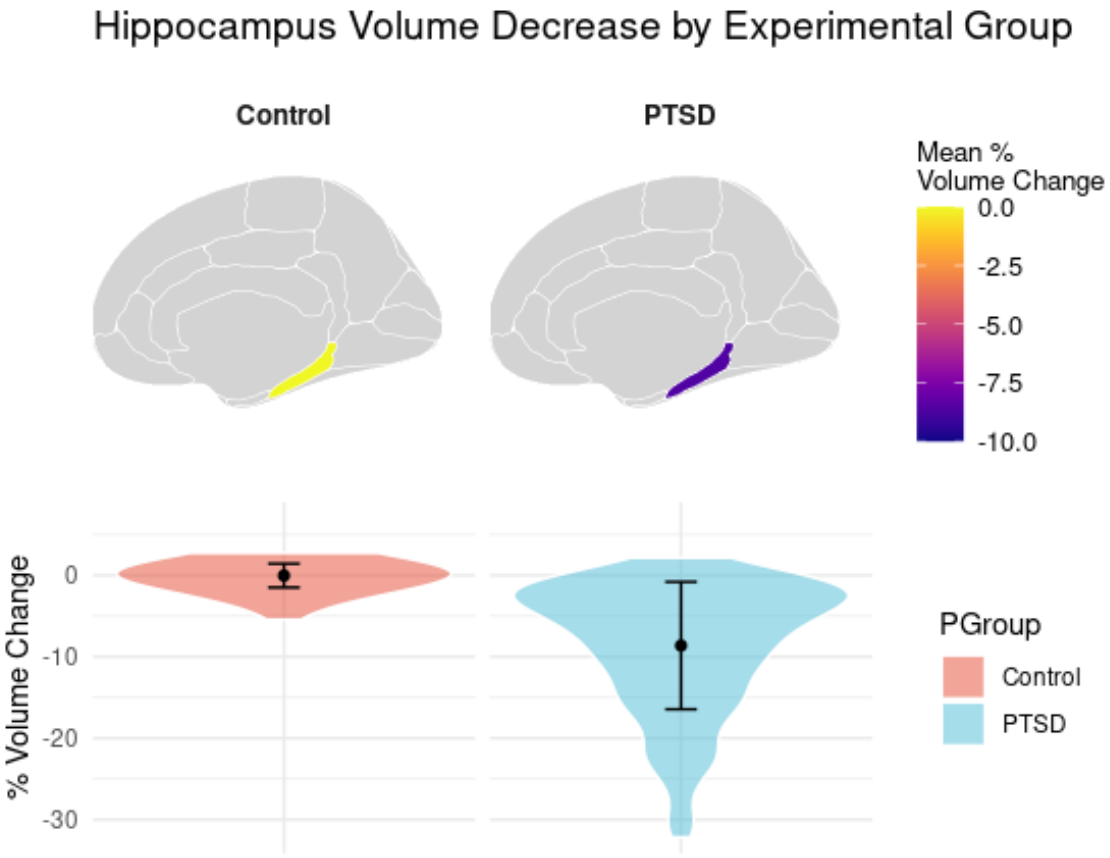

Figure 5.11. Predicted changes in hippocampal volume by experimental group. In the violin density plots (bottom), points represent means and error bars represent standard deviations.

The final analysis for hippocampal volume was whether or not the degree of hippocampal volume reduction was correlated to the degree of symptom severity. This is important because, although symptom severity is clearly driven by the severity of the traumatic event, it also depends on other factors that were explicitly manipulated in the simulations. To do so, the mean daily relative frequency of memory intrusions in the last 10 days of the simulations (red shaded

area in Figure 5.1) and the corresponding percentage decrease in hippocampal volume were calculated for each run of the model. Three separate linear regressions, one for each level of I = 20, 40, and 60, were then computed. In all cases, a significant linear regression was found [I = 20: β = -20.55, $t(7198)$ = -157.9, $p < .0001$; I = 40: β = -25.18, $t(7198)$ = -225.6, $p < .0001$; I = 60: β = -27.35, $t(7198)$ = 205.1, $p < .0001$] as shown in Figure 5.12.

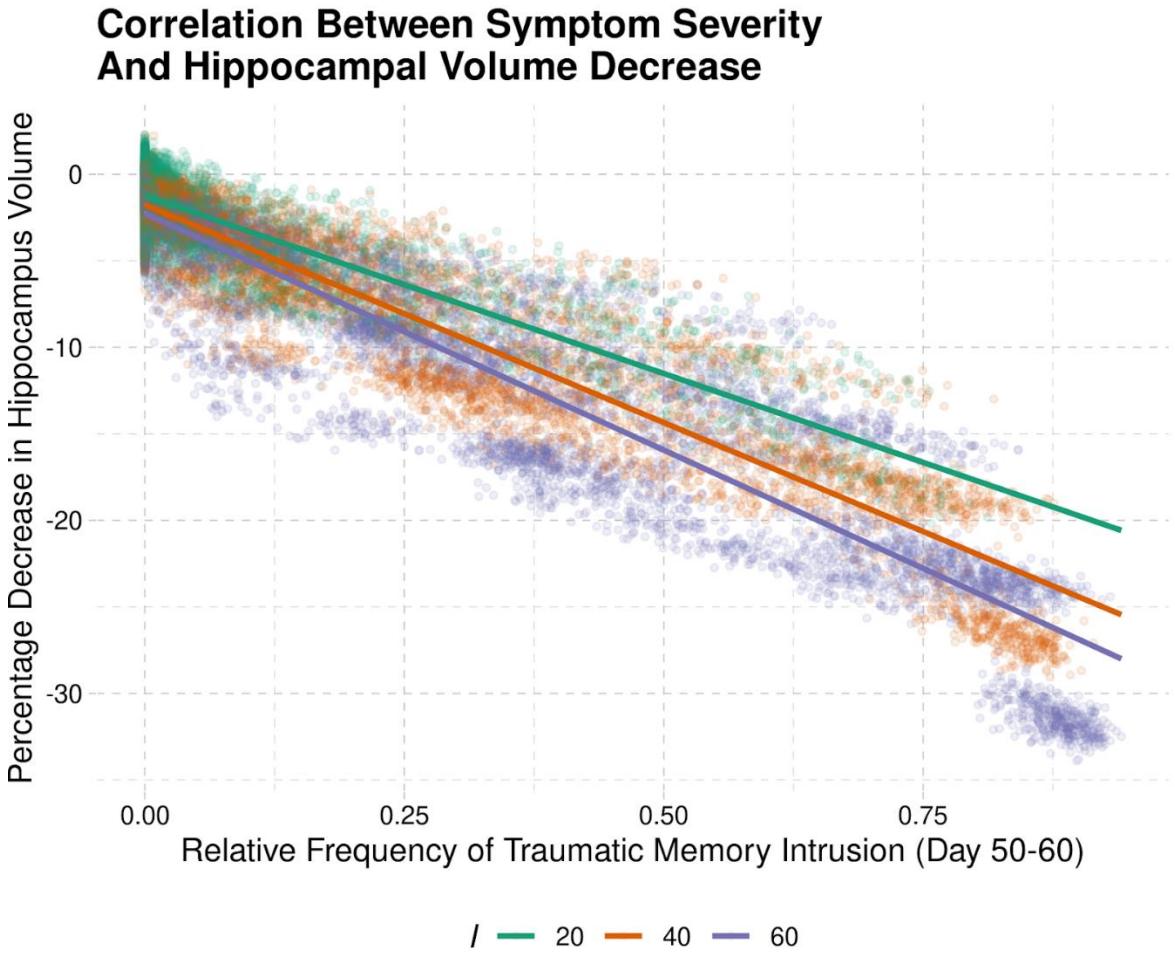

Figure 5.12. Correlation between the relative frequency of traumatic memory intrusions to hippocampal volume for varying levels of trauma intensity. Each point represents a single run of the model; solid lines represent the mean regression line. For the sake of clarity, all of the points corresponding to the control condition (I = 1) are omitted, since they are overlapping in the original of the coordinates (0, 0).

Because studies relating PTSD and hippocampal volume are correlational in nature, it is not possible to exclude the possibility that a smaller hippocampal volume might represent a risk factor for PTSD. This is, in fact, one of the most debated topics in the field, and some evidence in this sense can be found in the literature. In one of the most remarkable studies, Gilbertson et al 2002 [45] examined pairs of twins in which one of the sibling's suffered PTSD after being exposed to combat and found that the volume of the hippocampus in the non-exposed twin predicted the severity of the PTSD symptoms in the exposed one.

Although our model predicts that a decrease in hippocampal volume is a consequence of PTSD symptoms, it does allow for a potential way in which an *initially* smaller hippocampus could lead to worse clinical outcomes, thus explaining these results. In essence, our model predicts that the same set of conditions that would allow a traumatic event to dominate over all other memories could also lead to other uneven distribution of memory activations before any traumatic event. In other words, a smaller hippocampal volume could not be a risk factor *per se*, but evidence of the presence of other conditions (in this case, specific values of model parameters) that represent significant risk factors.

To examine this hypothesis, we conducted a new analysis of our simulations. In this analysis, as in the previous analysis, the hippocampal volume was measured as a percentage of the volume difference from a baseline condition in which no traumatic event occurs ($I = 1$). Unlike the previous analysis, however, the hippocampal volume was estimated from the average entropy in the ten days *preceding* the traumatic event (Days -11 to -1 in Figure 5.1). The percentage in volume difference from the baseline was then correlated with the mean severity of symptoms experienced by the model *after* the traumatic event, which was measured, like in the previous analysis, on Days 51-60.

To avoid any spurious effect in our analysis, two precautions were taken. First, the simulations in which the model was spontaneously ruminating over previous memories (U = 20, Table 4.2) *before* any traumatic event were excluded, since this condition would artificially warp the distribution of memory entropy and, thus, the predicted hippocampal volume. As a second precaution, the results of this simulation were analyzed separately for the different values of the cognitive control parameter, W (i.e., W = 4, 8, 12, Table 4.2). This was done to prevent Simpson's amalgamation paradox [123] in examining our data, since higher values of cognitive control are likely to increase entropy (and thus, predicted hippocampal volume) *before* a PTE, but to decrease entropy and PTSD symptoms *after* it.

The results of our analysis are presented in Figure 5.13. For each of the three cognitive control conditions, a significant negative linear effect was found (W = 4: β = -0.11, t(3598) = -23.0, $p < .0001$; W = 8: β = -0.03, t(3598) = -16.9, $p < .0001$; W = 12: β = -0.02, t(3598) = -11.7, $p < .0001$). Thus, our results confirm that, as hypothesized, conditions that make the model vulnerable to significant PTSD symptoms following a traumatic event do have a significant tendency to manifest themselves in the form of predicted smaller hippocampal volume. Furthermore, the values of the linear regression coefficients were significantly different across cognitive control conditions ($p < .0001$), suggesting that, as hypothesized, cognitive control plays a significant role in modulating the effect.

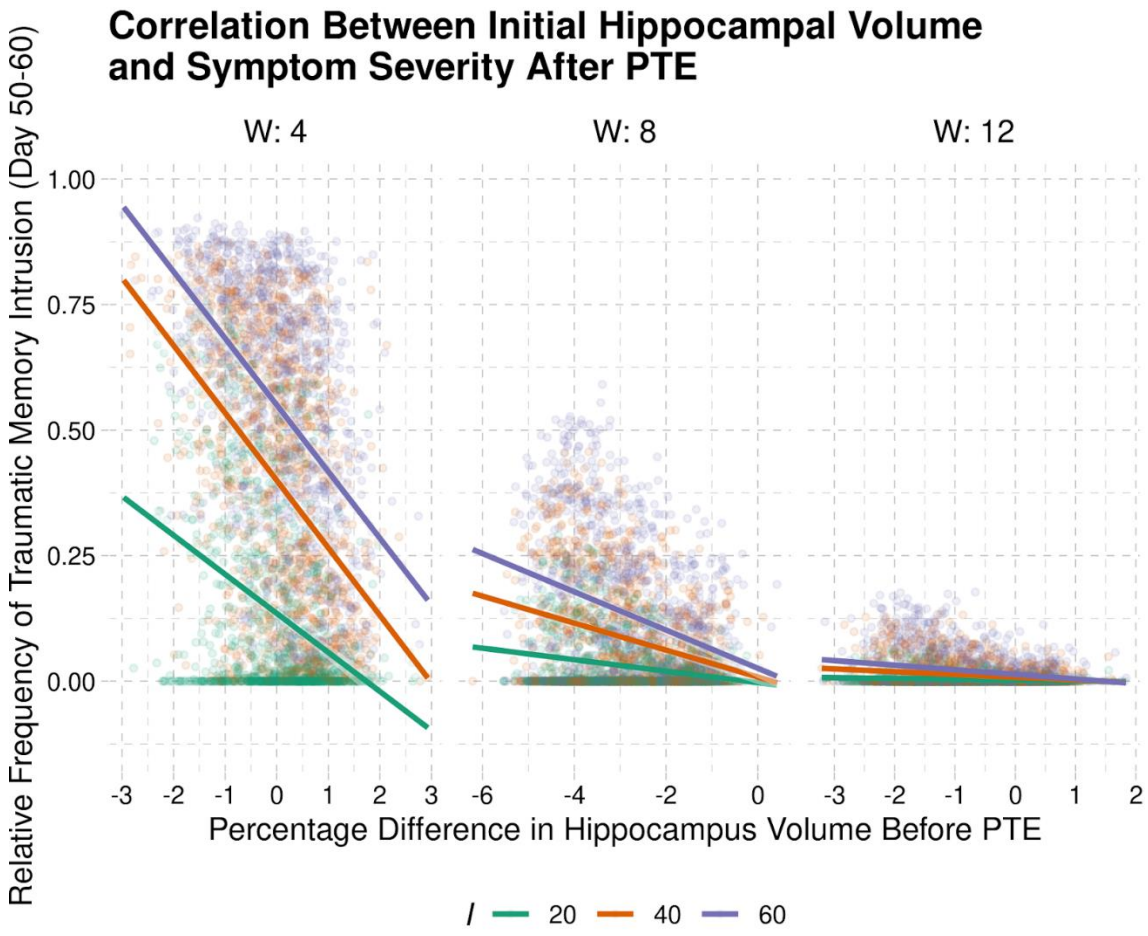

Figure 5.13. Hippocampal volume as a potential risk factor for PTSD. Changes in hippocampal volumes (measured as percentage differences from baseline) were negatively associated with the severity of PTSD symptoms following the PTE. As in Fig. 4, points corresponding to the initial baseline condition I =1 are concentrated on the axis origin (0, 0) and omitted for clarity.

5.5.2 *Functional Connectivity*

A similar analysis to Figure 5.11 was conducted for the predicted functional connectivity between the medial prefrontal cortex and the hippocampus. Functional connectivity was defined as the degree of similarity between retrieved memories and current context and is visualized in Figure 5.14. Because, in our simulations, the similarity between the traumatic memory and the context is experimentally manipulated as the factor C, this analysis was restricted to the subset of

simulations in which $C = 0$. The results showed that the simulated traumatized control group's runs (in which the model that were exposed to trauma but successfully recovered) exhibited significantly greater functional connectivity ($M = 0.53$, $SD = 0.20$) than the simulated PTSD group ($M = 0.12$, $SD = 0.15$; Welch two-sample test, $t(722.86) = 39.84$, $p < 0.0001$).

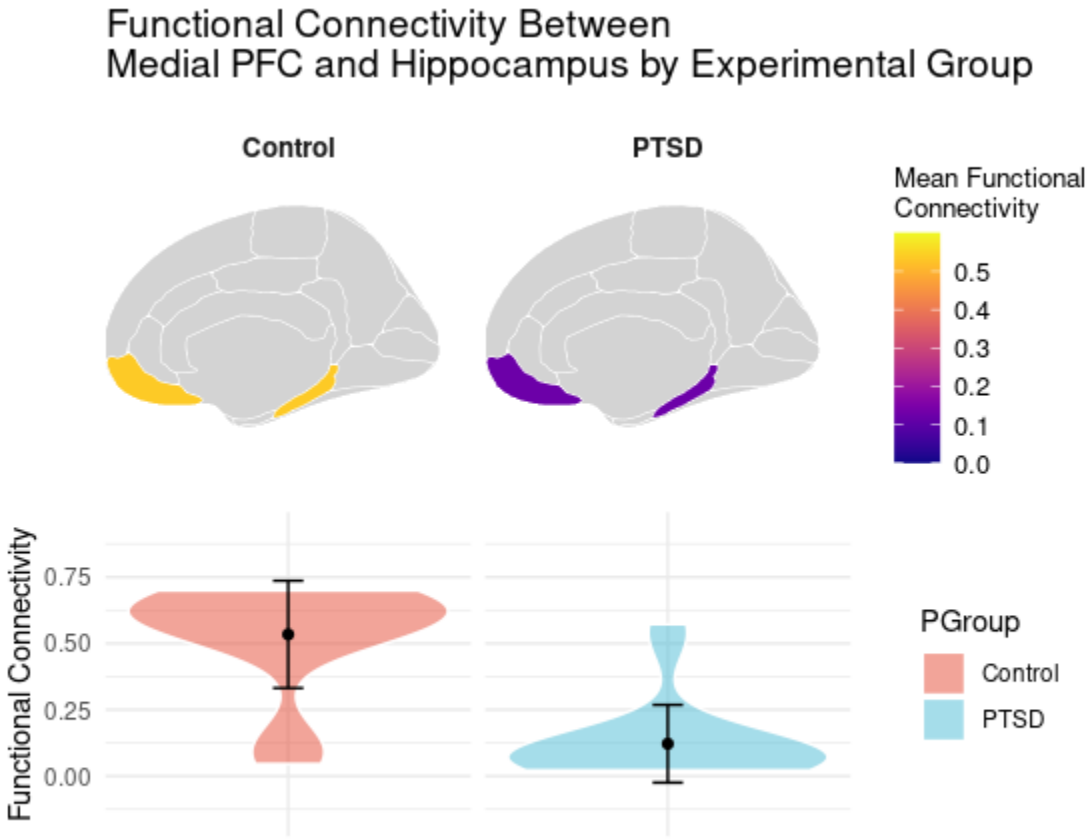

Figure 5.14. Predicted changes in functional connectivity between ventromedial PFC and hippocampus, divided by the simulated group of patients. In the violin density plots (bottom), points represent means and error bars represent standard deviations.

5.6 Fear Overgeneralization

Figure 5.15 illustrates the probability of a startle response, *P(startle)*, for three hypothetical levels of startle intensities $I(AX+)$ (i.e., ΔA, Eq. (4.11)), and at varying levels of log

$I(\neg m)$. Values for ΔA were selected to be proportional to the magnitude of the startle response observed in control, Low and High PTSD groups from Jovanovic et al. 2009, Figure 2 [48]. The increasing value of log $I(\neg m)$ represents the increases in intensity I of a traumatic memory, and/or an increase in the number of memories that have an elevated intensity value. The gray shaded areas represent the area of enhanced generalization for each condition. In this area, the probability of retrieving a shock stimulus and thus produce a startle response, is greater than what would be expected if no traumatic memory was present (i.e., when log $I(\neg m) = 0$).

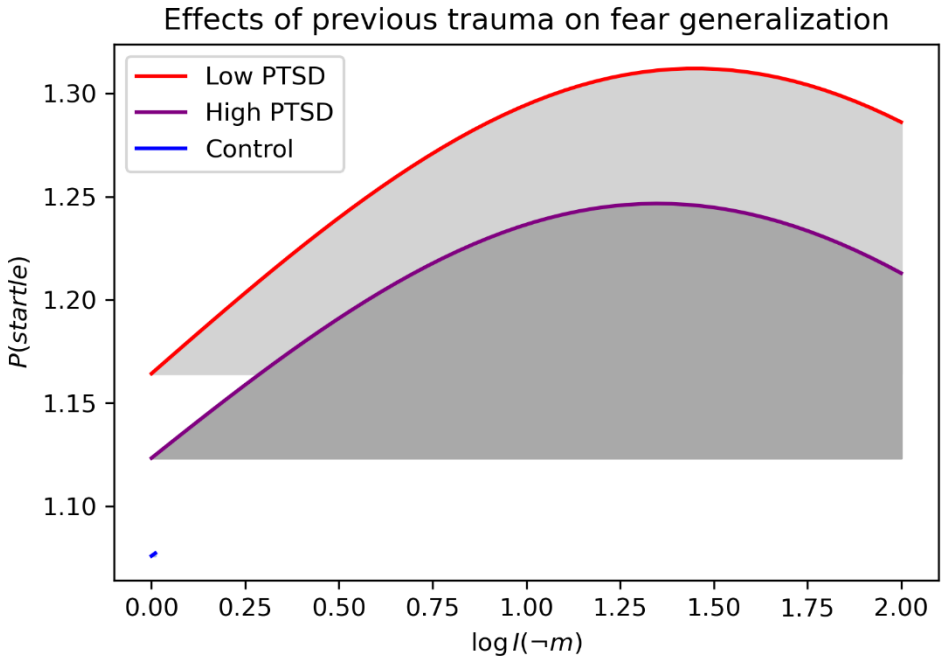

Figure 5.15. Theoretical startle probabilities for Low PTSD (red curve), High PTSD (purple curve) and control (blue point) conditions as a function of the average intensity of all non-task memories $I(\neg m)$.

For the Low and High PTSD conditions (red and purple curves, Figure 5.15), we see that as the value of log $I(\neg m)$ increases from 0 to about 1, there is an increase in the probability of retrieving a startle response as expected. For larger values of log $I(\neg m)$, however, we see the

counterintuitive result that the probability of a startle response levels off and begins decreasing. In our model, this is due to the simultaneous increase in the probability of retrieving a traumatic memory, instead of the task relevant memory, as log $I(\neg m)$ increases. In these circumstances, the traumatic memory 'intrudes' during the task in the presence of every stimulus. For the Control condition (blue point, Figure 5.15), log $I(\neg m) = 0$, and the value of *P(startle)* is 1.08.

The simulated results for the task presentation of a novel compound stimulus AB are shown in Figure 5.16. As predicted, the model produces an increase in the average percentage of startle responses for the presentation of the AB stimulus as the intensity and/or number of the traumatic memories increases (i.e., the average value of log $I(\neg m)$ increases).

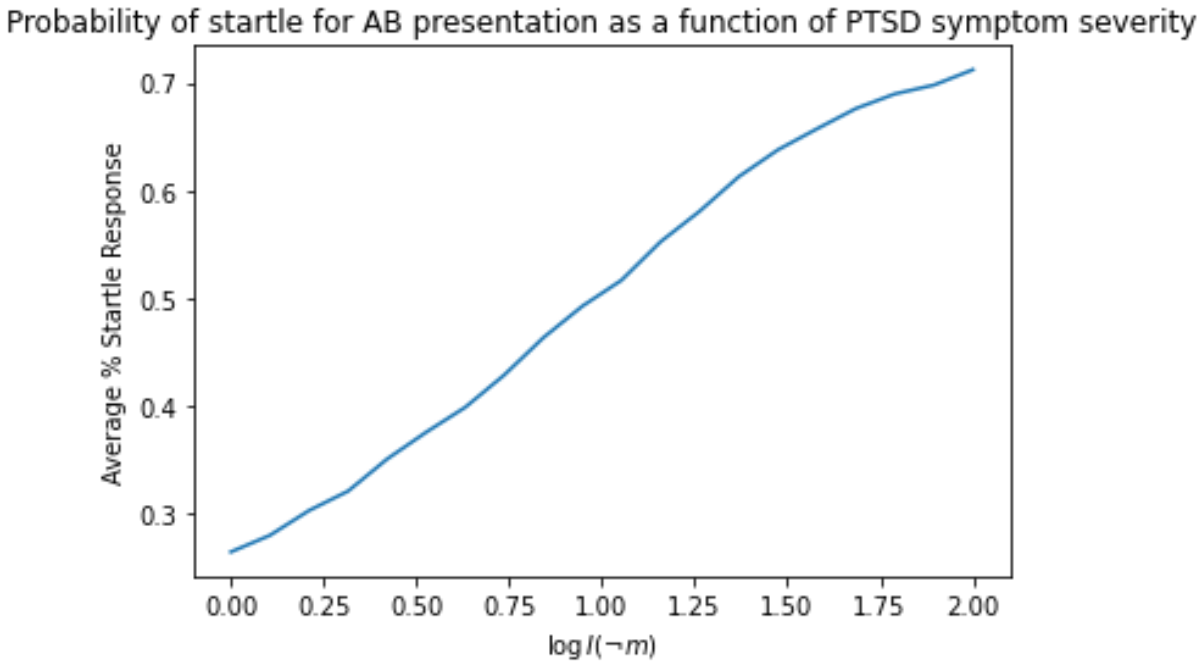

Figure 5.16. Average percent of startle responses that occurred in the model as a function of the PTSD symptom severity, i.e., the average intensity of all non-task memories *I(¬m)*.

The simulated results for the presentation of the AX, BX, AB, and noise alone (NA) conditions to control, low, and high PTSD symptom groups are shown in Figure 5.17. Values for

the control group are calculated with log $I(\neg m) = 0$, values for the low-symptom group are calculated with log $I(\neg m) = 0.65$, and values for the high-symptom group are calculated with log $I(\neg m) = 0.89$. As predicted, the model produces an increase in the average percentage of startle response for each condition as PTSD symptom severity increases.

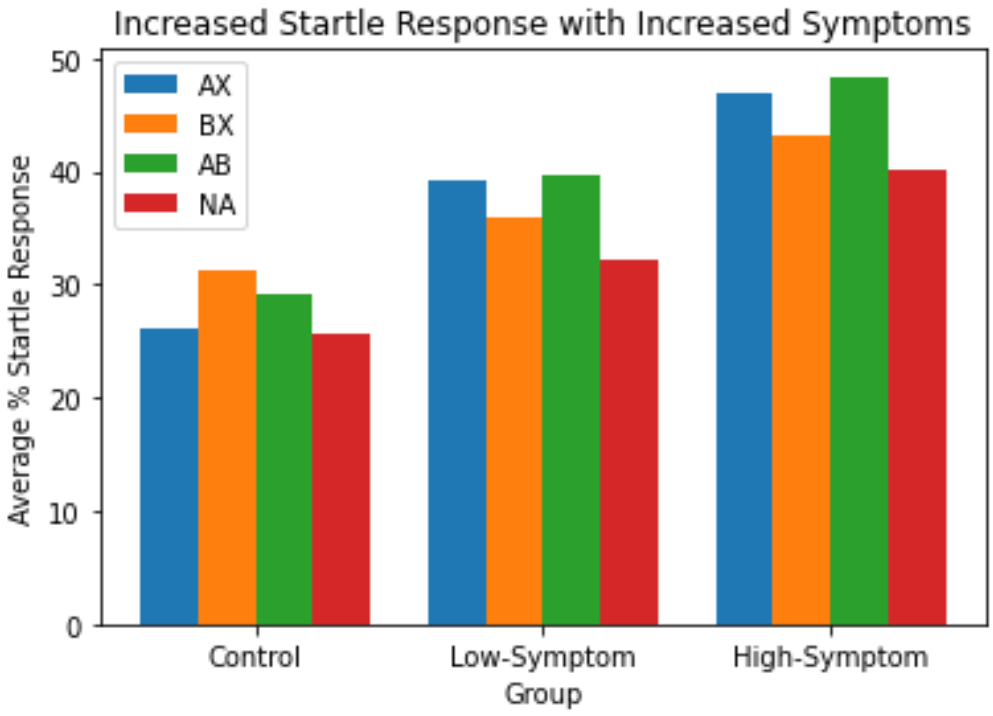

Figure 5.17. average percentage of startle response for a control group, and low and high symptom PTSD groups. Generally, we see that the average percentage of startle responses increases with symptom severity.

Figure 5.18 shows the percent potentiation of the startle response to the AX, BX, and AB conditions for each of the control, low, and high PTSD symptom groups. Percent potentiation was calculated as the startle response to AX, BX, or AB trials normalized to the NA condition for each group. Values for the control condition are calculated with log $I(\neg m) = 0$, values for the

low-symptom group are calculated with log $I(\neg m) = 0.65$, and values for the low-symptom group are calculated with log $I(\neg m) = 0.89$.

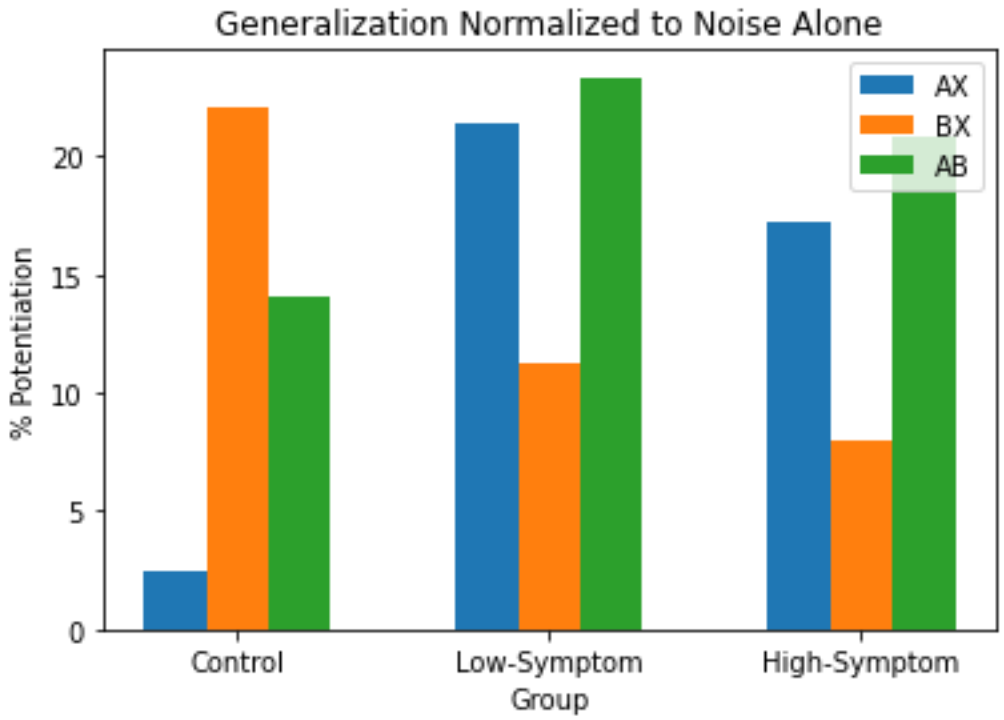

Figure 5.18. Normalized the startle response for AX, BX, and AB conditions to Noise alone to give the percent potentiation of the startle response to each condition.

Compared to what is seen in the literature [48] two key differences arise. First, the scale in Figure 5.18 is lower in magnitude, which is most likely due to the magnitude of the NA condition being higher than what is seen in the literature (Figure 5.17). Secondly, each group has an unexpected response to the AX condition. We expect to see AX potentiate a startle response more than all other conditions, even for controls, but in our model, this is not the case.

Chapter 6. DISCUSSION

In this thesis, I have described a computational model of intrusive memories following a traumatic event. The model is rooted in the rational analysis framework of human memory [4] and accounts for the insurgence of traumatic memories as a consequence of the additional activation boost that their emotional intensity confers. The model correctly predicts a number of behavioral findings, including the frequency of intrusive memories after the traumatic event and the effects that external and idiographic factors have on individual recovery trajectories. Additionally, the model also predicts a reduction in the size of the hippocampus and a reduction in the functional connectivity between prefrontal cortex and the hippocampus as a consequence of the persistence of intrusive memories. To the best of our knowledge, this is the first model to provide such a large coverage of neuroimaging and clinical findings.

The model predictions are noteworthy not only in extent but also in detail. The use of a realistic memory model and extended simulation times allow for detailed day prediction of the occurrence of intrusive memories under different contexts. In doing so, the work significantly tests the reach of the ACT-R declarative memory model to a timescale many orders of magnitudes larger than that of the simple laboratory tasks it was designed for. The fact the model could run and produce meaningful and realistic results on this extended timescale, even when the parameters were left unchanged (as was the case of the decay rate d), testifies to the robustness of this model. Such modeling over days, weeks, months, and years will be needed to extend this work to clinical applications.

6.1 LIMITATIONS

One of the main limitations of the model is the narrow way in which it captures learning. In our model, learning only occurs as a function of the general laws that govern memory decay over time, such as the power law of forgetting and the effects of recency and frequency. In contrast, previous models have focused on fear avoidance and learned cued associations [3], both of which were captured and expressed in terms of reinforcement and associative learning, respectively. Although our results show that the laws of memory decay and their modulation by emotion go a long way to capture a variety of behavioral and neurological phenomena, the effects of associative learning should not be discounted and certainly play an important role. For example, clinical psychotherapeutic interventions for PTSD are both successful and largely based on the associative tenets of fear learning and extinction [124–126]. Thus, these two approaches to modeling PTSD should not be taken as mutually exclusive, but as two different venues that focus on different facets of the disorder.

A related limitation is that our model addresses only intrusive memories per se but does not address the emotional and arousal states associated with intrusive memories, nor does it capture the behavioral changes that are associated with them. Intrusions themselves are not pathological and can differ on both frequency and distress. While frequency and distress are likely highly associated, it is possible to have intrusions without distress, with the latter not being pathological. Further, patients suffering from PTSD often report persistent anxiety and distress and modify their routines to actively avoid triggers. These changes have been previously addressed by the associative type of computational model [51] but are outside the reach of the model presented herein.

A word of caution should be also spent in comparing the model's behavioral predictions against the empirical data. This is because the samples involved in different studies are highly heterogeneous, varying significantly in the type of traumatic event and the time between the event and subsequent measurements. Our model captures the heterogeneity of the events only within a bidimensional space (intensity *I* and similarity to everyday context *C*), but it is likely that other dimensions might be needed to explain, for example, the differential effects of interpersonal trauma exposure and participation of sectarian violence [127]. Similarly, our model only captures the dynamics of recovery within a limited timeframe of two months, but different studies have reported data collected much beyond that horizon, and up to 20+ years after the traumatic event.

Another limitation of our model is that parameters are not allowed to vary as a function of trauma exposure or memory content. For example, the model currently assumes that an individual's cognitive control, or working memory capacity (captured by the parameter *W*), remains stable after the PTE, while experimental evidence suggests that it might be compromised by exposure to trauma [128].

Perhaps most remarkably, the parameter *d*, which controls the degree of memory decay and forgetting, was kept constant not only within but also across simulations. This choice was motivated by the fact that little is still known about individual differences in forgetting (although work quantifying such differences has recently been done: [73,74,129] and, as a consequence, not much is known about the interaction between forgetting and intrusive memories. Nonetheless, because of its critical role in determining the frequency of intrusive memories, future work should include careful examinations of the role of this parameter.

The most common explanation for the reduction of hippocampal volume is that prolonged stress leads to hippocampal damage, for example through the accumulation of stress-related hormones like cortisol. However, as noted by Garfinkel and Liberzon [130], cortisol levels are not universally or consistently elevated in individuals with PTSD; indeed, potentially even suppressed [131,132]. The model's estimates, instead, rely on a different logic, namely, that the hippocampus size reflects the landscape of memory availability and that, in PTSD, the dominant availability of the traumatic memory implicitly causes a reduced number of hippocampal cells and synapses devoted to storing memory information.

Finally, one last limitation that needs to be acknowledged is the duration of the simulations after the PTE. Because of their computational demands, our simulations only extended to 60 days after trauma. Clinical consensus suggests that trajectories would stabilize at a later time period (> 90 days). Because of this, our algorithm for trajectory classification takes into account the ascending or descending trend in the number of traumatic memory intrusions, rather than testing for a complete recovery at the endpoint. This fact might explain the paradoxical results of Figure 5.5, where the Recovery trajectory seems to increase, and the Delayed and Chronic trajectories decrease with the intensity I_{PTE} of the traumatic event. Thus, an extended simulation horizon in which the number of intrusive memories after three or more months would likely yield better and more reliable estimates of the trajectories.

6.2 Conclusions

These limitations notwithstanding, we believe that this computational model represents a first and important step into better understanding the nature of traumatic memories and their relationship to posttraumatic stress disorder and memory processing. Models such as this have the ability to manipulate parameters identified in the clinical literature and link them to

behavioral and neurobiological effects, showing and confirming functional relationships that simply are not possible to manipulate in the real world. Critically, they have the ability to capture and model dynamic retrieval processes characterized by episodic event processing. Indeed, strong predictors emerged. Both higher traumatic event severity and congruence with the current environment resulted in higher intrusions of traumatic memories. Similarly, lower cognitive control and higher post-event unproductive processing and rumination also resulted in more intrusive memories. These results converge with the clinical literature and their emergence from a functional computational model of memory argues beyond simply Granger causality, moving the field towards the understanding these factors likely have causal properties in psychiatric disorders.

Future studies using ACT-REM will help to determine the strength of causal links between environmental and idiographic factors and clinical mental health outcomes. Objective cognitive tasks and assessments can be used to determine individual values for model parameters such as environmental congruency (did the PTE occur in a novel or familiar environment?), cognitive control and executive functioning, memory decay rate, PTE intensity and vividness of re-experiencing, and frequency of unproductive processing. These features can then be monitored over time in relation to treatment progression and symptom severity. Furthermore, any one of these model parameters could be improved or changed in patients to assess their specific impact on PTSD symptoms and recovery, and to further validate the model. Targeted neuromodulation, via neurostimulation techniques such as transcranial magnetic stimulation (TMS) or deep brain stimulation (DBS), offers researchers the opportunity to isolate individual brain regions and their cognitive functions. This sort of targeted exploration will help to determine whether improving one cognitive feature is more effective than others in aiding

recovery from PTSD. Specifically, I expect that improvement in cognitive control, via training tasks or targeted neuromodulation, would have the largest impact and result not only in improved PTSD outcomes, but simultaneous improvement in other parameters such as memory decay rate, vividness of re-experiencing, and frequency of unproductive processing.

Finally, this model might have relevant impacts in clinical mental health treatment. Despite remarkable advances, the field of computational psychiatry [60,133,134] has fallen short of the possibility of delivering individually parametrized models that could be realistically used to aid in treating patients. However, the degree of detail and realism achieved by the simulations presented in this thesis, in both their timescale and the detail of behavioral and neural predictions, make it possible to consider using this model in realistic mental health applications. For example, this model could be calibrated on individual patient data and provide mobile health recommendation, data collection, and monitoring [135,136]. This patient specific information would be a game changer for disorders like PTSD. In the immediate aftermath of a potentially traumatic event, patients who are most likely to have chronic or severe psychopathology can be identified and triaged, thereby reducing the time it takes to receive appropriate psychological intervention and improving treatment outcomes. This sort of computational psychiatry is an exciting avenue for future research because early and individualized assessment for PTSD has been elusive in the field and barriers to receiving early intervention are high, despite how debilitating symptoms of PTSD can be [137].

BIBLIOGRAPHY

1. Hagenaars MA, van Minnen A, Hoogduin CAL, Verbraak M. A transdiagnostic comparison of trauma and panic memories in PTSD, panic disorder, and healthy controls. J Behav Ther Exp Psychiatry. 2009;40: 412–422.

2. Bywaters M, Andrade J, Turpin G. Intrusive and non-intrusive memories in a non-clinical sample: The effects of mood and affect on imagery vividness. Memory. 2004;12: 467–478.

3. Radell ML, Myers CE, Sheynin J. Computational models of post-traumatic stress disorder (PTSD). Computational models. 2017. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119159193#page=67

4. Anderson JR. The Adaptive Character of Thought. Mahwah, NJ: Lawrence Erlbaum Associates; 1990.

5. Shin LM, Rauch SL, Pitman RK. Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. Ann N Y Acad Sci. 2006;1071: 67–79.

6. Liberzon I, Sripada CS. The functional neuroanatomy of PTSD: a critical review. Prog Brain Res. 2008;167: 151–169.

7. Marks EH, Franklin AR, Zoellner LA. Can't get it out of my mind: A systematic review of predictors of intrusive memories of distressing events. Psychol Bull. 2018. Available: https://psycnet.apa.org/record/2018-11360-001

8. O'Doherty DCM, Chitty KM, Saddiqui S, Bennett MR, Lagopoulos J. A systematic review and meta-analysis of magnetic resonance imaging measurement of structural volumes in posttraumatic stress disorder. Psychiatry Res. 2015;232: 1–33.

9. Asok A, Kandel ER, Rayman JB. The Neurobiology of Fear Generalization. Front Behav Neurosci. 2018;12: 329.

10. Michael T, Ehlers A, Halligan SL, Clark DM. Unwanted memories of assault: what intrusion characteristics are associated with PTSD? Behav Res Ther. 2005;43: 613–628.

11. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. Clin Psychol Rev. 2018;63: 41–55.

12. Galatzer-Levy IR, Bonanno GA, Bush DEA, Ledoux JE. Heterogeneity in threat extinction learning: substantive and methodological considerations for identifying individual difference in response to stress. Front Behav Neurosci. 2013;7: 55.

13. Frans O, Rimmö P-A, Aberg L, Fredrikson M. Trauma exposure and post-traumatic stress disorder in the general population. Acta Psychiatr Scand. 2005;111: 291–299.

14. Kolassa I-T, Ertl V, Eckart C, Kolassa S, Onyut LP, Elbert T. Spontaneous remission from PTSD depends on the number of traumatic event types experienced. Psychol Trauma. 2010;2: 169–174.

15. Besnard A, Sahay A. Adult Hippocampal Neurogenesis, Fear Generalization, and Stress. Neuropsychopharmacology. 2016;41: 24–44.

16. Peverill M, Finn AS, McLaughlin KA. Dimensions of childhood adversity have distinct associations with neural systems underlying executive functioning. Development. 2017. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/pmc5733141/

17. Cohodes EM, Chen SH, Lieberman AF, Bush NR. Examination of the associations between young children's trauma exposure, trauma-symptomatology, and executive function. Child Abuse Negl. 2020;108: 104635.

18. Bomyea J, J. Lang A. The Role of Executive Functioning in PTSD and its Treatment. Curr Psychiatry Rev. 2015;11: 160–171.

19. van Rooij SJH, Jovanovic T. Impaired inhibition as an intermediate phenotype for PTSD risk and treatment response. Prog Neuropsychopharmacol Biol Psychiatry. 2019;89: 435–445.

20. van Rooij SJH, Rademaker AR, Kennis M, Vink M, Kahn RS, Geuze E. Impaired right inferior frontal gyrus response to contextual cues in male veterans with PTSD during response inhibition. J Psychiatry Neurosci. 2014;39: 330–338.

21. Weber DL, Clark CR, McFarlane AC, Moores KA, Morris P, Egan GF. Abnormal frontal and parietal activity during working memory updating in post-traumatic stress disorder. Psychiatry Res. 2005;140: 27–44.

22. Saunders N, Downham R, Turman B, Kropotov J, Clark R, Yumash R, et al. Working memory training with tDCS improves behavioral and neurophysiological symptoms in pilot group with post-traumatic stress disorder (PTSD) and with poor working memory. Neurocase. 2015;21: 271–278.

23. Roxanne Sopp M, Streb M, Brueckner AH, Schäfer SK, Lass-Hennemann J, Mecklinger A, et al. Prospective associations between intelligence, working memory capacity, and intrusive memories of a traumatic film: Potential mediating effects of rumination and memory disorganization. J Behav Ther Exp Psychiatry. 2021;70: 101611.

24. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. Cogn Psychol. 2000;41: 49–100.

25. Berntsen D, Willert M, Rubin DC. Splintered memories or vivid landmarks? Qualities and organization of traumatic memories with and without PTSD. Appl Cogn Psychol. 2003;17: 675–693.

26. Morina N, Leibold E, Ehring T. Vividness of general mental imagery is associated with the occurrence of intrusive memories. J Behav Ther Exp Psychiatry. 2013;44: 221–226.

27. Hackmann A, Ehlers A, Speckens A, Clark DM. Characteristics and content of intrusive memories in PTSD and their changes with treatment. J Trauma Stress. 2004;17: 231–240.

28. Akiki TJ, Averill CL, Abdallah CG. A Network-Based Neurobiological Model of PTSD: Evidence From Structural and Functional Neuroimaging Studies. Curr Psychiatry Rep. 2017;19: 81.

29. Koenigs M, Grafman J. Posttraumatic stress disorder: the role of medial prefrontal cortex and amygdala. Neuroscientist. 2009;15: 540–548.

30. Shin LM, Wright CI, Cannistraro PA, Wedig MM, McMullin K, Martis B, et al. A functional magnetic resonance imaging study of amygdala and medial prefrontal cortex responses to overtly presented fearful faces in posttraumatic stress disorder. Arch Gen Psychiatry. 2005;62: 273–281.

31. Shin LM, Orr SP, Carson MA, Rauch SL, Macklin ML, Lasko NB, et al. Regional Cerebral Blood Flow in the Amygdala and Medial PrefrontalCortex During Traumatic Imagery in Male and Female Vietnam Veterans With PTSD. Arch Gen Psychiatry. 2004;61: 168–176.

32. Nadel L, Samsonovich A, Ryan L, Moscovitch M. Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. Hippocampus. 2000;10: 352–368.

33. Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, et al. Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. J Anat. 2005;207: 35–66.

34. Nadel L, Moscovitch M. Memory consolidation, retrograde amnesia and the hippocampal complex. Curr Opin Neurobiol. 1997;7: 217–227.

35. McGaugh JL. Memory--a century of consolidation. Science. 2000;287: 248–251.

36. Alvarez P, Squire LR. Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci U S A. 1994;91: 7041–7045.

37. Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. Nat Rev Neurosci. 2012;13: 713–726.

38. Bremner JD, Randall P, Vermetten E, Staib L, Bronen RA, Mazure C, et al. Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse--a preliminary report. Biol Psychiatry. 1997;41: 23–32.

39. Bremner JD, Randall P, Scott TM, Bronen RA, Seibyl JP, Southwick SM, et al. MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. Am J Psychiatry. 1995;152: 973–981.

40. Villarreal G, Hamilton DA, Petropoulos H, Driscoll I, Rowland LM, Griego JA, et al. Reduced hippocampal volume and total white matter volume in posttraumatic stress disorder. Biol Psychiatry. 2002;52: 119–125.

41. Nelson MD, Tumpap AM. Posttraumatic stress disorder symptom severity is associated with left hippocampal volume reduction: a meta-analytic study. CNS Spectr. 2017;22: 363–372.

42. Logue MW, van Rooij SJH, Dennis EL, Davis SL, Hayes JP, Stevens JS, et al. Smaller hippocampal volume in posttraumatic stress disorder: a multisite ENIGMA-PGC study: subcortical volumetry results from posttraumatic stress disorder consortia. Biol Psychiatry. 2018;83: 244–253.

43. Karl A, Schaefer M, Malta LS, Dörfel D, Rohleder N, Werner A. A meta-analysis of structural brain abnormalities in PTSD. Neurosci Biobehav Rev. 2006;30: 1004–1031.

44. Pitman RK, Gilbertson MW, Gurvits TV, May FS, Lasko NB, Metzger LJ, et al. Clarifying the origin of biological abnormalities in PTSD through the study of identical twins discordant for combat exposure. Ann N Y Acad Sci. 2006;1071: 242–254.

45. Gilbertson MW, Shenton ME, Ciszewski A, Kasai K, Lasko NB, Orr SP, et al. Smaller hippocampal volume predicts pathologic vulnerability to psychological trauma. Nat Neurosci. 2002;5: 1242–1247.

46. Grillon C, Morgan CA 3rd. Fear-potentiated startle conditioning to explicit and contextual cues in Gulf War veterans with posttraumatic stress disorder. J Abnorm Psychol. 1999;108: 134–142.

47. Jovanovic T, Norrholm SD, Blanding NQ, Davis M, Duncan E, Bradley B, et al. Impaired fear inhibition is a biomarker of PTSD but not depression. Depress Anxiety. 2010;27: 244–251.

48. Jovanovic T, Norrholm SD, Fennell JE, Keyes M, Fiallos AM, Myers KM, et al. Posttraumatic stress disorder may be associated with impaired fear inhibition: relation to symptom severity. Psychiatry Res. 2009;167: 151–160.

49. Morey RA, Haswell CC, Stjepanović D, Mid-Atlantic MIRECC Workgroup, Dunsmoor JE, LaBar KS. Neural correlates of conceptual-level fear generalization in posttraumatic stress disorder. Neuropsychopharmacology. 2020;45: 1380–1389.

50. Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press Cambridge; 1998.

51. Radell ML, Beck KD, Pang KCH, Myers CE. Using signals associated with safety in avoidance learning: computational model of sex differences. PeerJ. 2015;3: e1081.

52. Myers CE, Moustafa AA, Sheynin J, Vanmeenen KM, Gilbertson MW, Orr SP, et al. Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: empirical data and computational model. PLoS One. 2013;8: e72508.

53. Radell ML, Sanchez R, Weinflash N, Myers CE. The personality trait of behavioral inhibition modulates perceptions of moral character and performance during the trust game: behavioral results and computational modeling. PeerJ. 2016;4: e1631.

54. Sheynin J, Moustafa AA, Beck KD, Servatius RJ, Myers CE. Testing the role of reward and punishment sensitivity in avoidance behavior: a computational modeling approach. Behav Brain Res. 2015;283: 121–138.

55. Rubin DC, Berntsen D, Bohni MK. A memory-based model of posttraumatic stress disorder: evaluating basic assumptions underlying the PTSD diagnosis. Psychol Rev. 2008;115: 985–1011.

56. Cohen RT, Kahana MJ. A memory-based theory of emotional disorders. Psychol Rev. 2022;129: 742–776.

57. White CN, Curl RA, Sloane JF. Using Decision Models to Enhance Investigations of Individual Differences in Cognitive Neuroscience. Front Psychol. 2016;7: 81.

58. White CN, Congdon E, Mumford JA, Karlsgodt KH, Sabb FW, Freimer NB, et al. Decomposing decision components in the stop-signal task: a model-based approach to individual differences in inhibitory control. J Cogn Neurosci. 2014;26: 1601–1614.

59. Redish AD. Addiction as a computational process gone awry. Science. 2004;306: 1944–1947.

60. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19: 404–413.

61. Anderson JR. How Can the Mind Occur in the Physical Universe? Oxford University Press; 2007.

62. Kotseruba I, Tsotsos JK. 40 years of cognitive architectures: core cognitive abilities and practical applications. Artificial Intelligence Review. 2020;53: 17–94.

63. Fum D, Stocco A. Memory, Emotion, and Rationality: An ACT-R interpretation for Gambling Task results. In: C. D. Schunn, M. C. Lovett, C. Lebiere & P. Munro, editor. Proceedings of the 6th International Conference on Cognitive Modeling. Lawrence Erlbaum Associates; 2004. pp. 106–111.

64. Dancy CL. ACT-RΦ: A cognitive architecture with physiology and affect. Biologically Inspired Cognitive Architectures. 2013;6: 40–45.

65. Juvina I, Larue O, Hough A. Modeling valuation and core affect in a cognitive architecture: The impact of valence and arousal on memory and decision-making. Cogn Syst Res. 2018;48: 4–24.

66. Anderson JR, Schooler LJ. Reflections of the environment in memory. Psychol Sci. 1991;2: 396–408.

67. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. Psychol Rev. 2004;111: 1036–1060.

68. Anderson JR. How Can the Human Mind Occur in the Physical Universe? Oxford University Press; 2009.

69. Pavlik PI Jr, Anderson JR. Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. Cogn Sci. 2005;29: 559–586.

70. Anderson JR. Retrieval of propositional information from long-term memory. Cogn Psychol. 1974;6: 451–474.

71. Anderson JR, Bothell D, Lebiere C, Matessa M. An Integrated Theory of List Memory. J Mem Lang. 1998/5;38: 341–380.

72. Pavlik PI, Anderson JR. Using a model to compute the optimal schedule of practice. J Exp Psychol Appl. 2008;14: 101–117.

73. Sense F, Behrens F, Meijer RR, van Rijn H. An individual's rate of forgetting is stable over time but differs across materials. Top Cogn Sci. 2016;8: 305–321.

74. Sense F, Meijer RR, van Rijn H. Exploration of the Rate of Forgetting as a Domain-Specific Individual Differences Measure. Frontiers in Education. 2018;3: 112.

75. Anderson JR. A spreading activation theory of memory. Journal of verbal learning and verbal behavior. 1983. Available: https://www.sciencedirect.com/science/article/pii/S0022537183902013

76. Roelofs A. A spreading-activation theory of lemma retrieval in speaking. Cognition. 1992;42: 107–142.

77. Roediger HL III, Balota DA, Watson JM. Spreading activation and arousal of false memories. 2001. Available: https://psycnet.apa.org/record/2001-00138-006

78. Collins AM, Loftus EF. A spreading-activation theory of semantic processing. Psychol Rev. 1975;82: 407.

79. Kane MJ, Engle RW. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. Psychon Bull Rev. 2002;9: 637–671.

80. Kane MJ, Engle RW. Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. J Exp Psychol Gen. 2003;132: 47–70.

81. Burgess GC, Gray JR, Conway ARA, Braver TS. Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. J Exp Psychol Gen. 2011;140: 674–692.

82. Baddeley AD, Logie RH. Working memory: The multiple-component model. 1999. Available: https://psycnet.apa.org/record/1999-02490-001

83. Baddeley A. Working memory. Curr Biol. 2010;20: R136–R140.

84. Lovett MC, Daily LZ, Reder LM. A source activation theory of working memory: cross-task prediction of performance in ACT-R. Cogn Syst Res. 2000;1: 99–118.

85. Daily LZ, Lovett MC, Reder LM. Modeling individual differences in working memory performance: a source activation account. Cogn Sci. 2001;25: 315–353.

86. Rice, P. J., & Stocco, A. Estimating individual differences in working memory through ACT-R modeling and resting state connectivity. In: Stewart TC, editor. Proceedings of the 19th International Conference on Cognitive Modelling. University Park, PA: Penn State; 2021. pp. 241–247.

87. Murty VP, Tompary A, Adcock RA, Davachi L. Selectivity in Postencoding Connectivity with High-Level Visual Cortex Is Associated with Reward-Motivated Memory. J Neurosci. 2017;37: 537–545.

88. Murty VP, FeldmanHall O, Hunter LE, Phelps EA, Davachi L. Episodic memories predict adaptive value-based decision-making. J Exp Psychol Gen. 2016;145: 548–558.

89. Patil A, Murty VP, Dunsmoor JE, Phelps EA, Davachi L. Reward retroactively enhances memory consolidation for related items. Learn Mem. 2017;24: 65–69.

90. Cahill L, Gorski L, Le K. Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. Learn Mem. 2003;10: 270–274.

91. Kensinger EA, Schacter DL. Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. J Neurosci. 2006;26: 2564–2570.

92. Paul J. H. Schoemaker. The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. J Econ Lit. 1982;20: 529–563.

93. Panksepp J. Affective Neuroscience: The Foundations of Human and Animal Emotions. Oxford University Press; 2004.

94. Ledoux J. The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon and Schuster; 1998.

95. LeDoux J. Rethinking the emotional brain. Neuron. 2012;73: 653–676.

96. Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol. 2005;17: 715–734.

97. Feldman Barrett L, Russell JA. Independence and bipolarity in the structure of current affect. J Pers Soc Psychol. 1998;74: 967–984.

98. Bryant RA, Kemp AH, Felmingham KL, Liddell B, Olivieri G, Peduto A, et al. Enhanced amygdala and medial prefrontal activation during nonconscious processing of fear in posttraumatic stress disorder: an fMRI study. Hum Brain Mapp. 2008;29: 517–523.

99. Rolls and Alessandro Treves E. Neural Networks and Brain Function. OXFORD UNIVERSITY PRESS; 1998.

100. McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev. 1995;102: 419.

101. Hardt O, Nader K, Nadel L. Decay happens: the role of active forgetting in memory. Trends Cogn Sci. 2013;17: 111–120.

102. Maguire EA, Woollett K, Spiers HJ. London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus. 2006;16: 1091–1101.

103. Noble KG, Grieve SM, Korgaonkar MS, Engelhardt LE, Griffith EY, Williams LM, et al. Hippocampal volume varies with educational attainment across the life-span. Front Hum Neurosci. 2012;6: 307.

104. Woollett K, Maguire EA. Acquiring "the Knowledge" of London's layout drives structural brain changes. Curr Biol. 2011;21: 2109–2114.

105. Huffman DA. A Method for the Construction of Minimum-Redundancy Codes. Proceedings of the IRE. 1952;40: 1098–1101.

106. Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;27: 379–423.

107. Danker JF, Anderson JR. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol Bull. 2010;136: 87–102.

108. Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci U S A. 1997;94: 14792–14797.

109. Danker JF, Gunn P, Anderson JR. A rational account of memory predicts left prefrontal activation during controlled retrieval. Cereb Cortex. 2008;18: 2674–2685.

110. Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007;45: 2883–2901.

111. Stocco A, Anderson JR. Endogenous control and task representation: an fMRI study in algebraic problem-solving. J Cogn Neurosci. 2008;20: 1300–1314.

112. Anderson JR, Fincham JM, Qin Y, Stocco A. A central circuit of the mind. Trends Cogn Sci. 2008;12: 136–143.

113. Collins AGE. The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. J Cogn Neurosci. 2018;30: 1422–1432.

114. Tange O. GNU Parallel-the command-line power tool. The USENIX Magazine. 2011;36: 42–47.

115. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: https://www.R-project.org/

116. Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3: 180–185.

117. Mowinckel AM, Vidal-Piñeiro D. Visualization of Brain Statistics With R Packages ggseg and ggseg3d. Advances in Methods and Practices in Psychological Science. 2020;3: 466–483.

118. Jovanovic T, Keyes M, Fiallos A, Myers KM, Davis M, Duncan EJ. Fear potentiation and fear inhibition in a human fear-potentiated startle paradigm. Biol Psychiatry. 2005;57: 1559–1564.

119. Whitman JB, North CS, Downs DL, Spitznagel EL. A prospective study of the onset of PTSD symptoms in the first month after trauma exposure. Ann Clin Psychiatry. 2013;25: 163–172.

120. Kaldewaij R, Koch SBJ, Hashemi MM, Zhang W, Klumpers F, Roelofs K. Anterior prefrontal brain activity during emotion control predicts resilience to post-traumatic stress symptoms. Nat Hum Behav. 2021. doi:10.1038/s41562-021-01055-2

121. Mevissen L, de Jongh A. PTSD and its treatment in people with intellectual disabilities: a review of the literature. Clin Psychol Rev. 2010;30: 308–316.

122. Smith ME. Bilateral hippocampal volume reduction in adults with post-traumatic stress disorder: a meta-analysis of structural MRI studies. Hippocampus. 2005;15: 798–807.

123. Wagner CH. Simpson's Paradox in Real Life. Am Stat. 1982;36: 46–48.

124. Cusack K, Jonas DE, Forneris CA, Wines C, Sonis J, Middleton JC, et al. Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. Clin Psychol Rev. 2016;43: 128–141.

125. Graham BM, Milad MR. The study of fear extinction: implications for anxiety disorders. Am J Psychiatry. 2011;168: 1255–1265.

126. Lebois LAM, Seligowski AV, Wolff JD, Hill SB, Ressler KJ. Augmentation of Extinction and Inhibitory Learning in Anxiety and Trauma-Related Disorders. Annu Rev Clin Psychol. 2019;15: 257–284.

127. Kessler RC, Aguilar-Gaxiola S, Alonso J, Benjet C, Bromet EJ, Cardoso G, et al. Trauma and PTSD in the WHO World Mental Health Surveys. Eur J Psychotraumatol. 2017;8: 1353383.

128. El-Hage W, Gaillard P, Isingrini M, Belzung C. Trauma-related deficits in working memory. Cogn Neuropsychiatry. 2006;11: 33–46.

129. Zhou P, Sense F, van Rijn H, Stocco A. Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. Cognition. 2021;212: 104660.

130. Garfinkel Sarah N., Liberzon Israel. Neurobiology of PTSD: A Review of Neuroimaging Findings. Psychiatr Ann. 2009;39. doi:10.3928/00485713-20090527-01

131. Pan X, Wang Z, Wu X, Wen SW, Liu A. Salivary cortisol in post-traumatic stress disorder: a systematic review and meta-analysis. BMC Psychiatry. 2018;18: 324.

132. Morris MC, Hellman N, Abelson JL, Rao U. Cortisol, heart rate, and blood pressure as early markers of PTSD risk: A systematic review and meta-analysis. Clin Psychol Rev. 2016;49: 79–91.

133. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci. 2012;16: 72–80.

134. Wiecki TV, Poland J, Frank MJ. Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification. Clin Psychol Sci. 2015;3: 378–399.

135. Springer A, Venkatakrishnan A, Mohan S, Nelson L, Silva M, Pirolli P. Leveraging self-affirmation to improve behavior change: a mobile health app experiment. JMIR mHealth and uHealth. 2018;6: e157.

136. Pirolli P, Mohan S, Venkatakrishnan A, Nelson L, Silva M, Springer A. Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model. J Med Internet Res. 2017;19: e397.

137. Shalev AY, Ankri YLE, Peleg T, Israeli-Shalev Y, Freedman S. Barriers to receiving early care for PTSD: results from the Jerusalem trauma outreach and prevention study. Psychiatr Serv. 2011;62: 765–773.