Predicting the future to control the past : predictive computations underlying the brain dynamics of memory control and their role in understanding post-traumatic stress disorder evolution following the 13th November terrorist attacks

Giovanni Leone

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité PSYCHOLOGIE

Préparée au sein de l'Université de Caen Normandie

Predicting the future to control the past: predictive computations underlying the brain dynamics of memory control and their role in understanding post-traumatic stress disorder evolution following the 13th November terrorist attacks

Présentée et soutenue par
GIOVANNI LEONE

Thèse soutenue le 13/12/2021 devant le jury composé de		
MME PEGGY SERIES	Chercheur HDR, Université Edinbourg	Rapporteur du jury
M. FABIEN VINCKIER	Professeur praticien hospitalier, Institut du Cerveau	Rapporteur du jury
M. STEPHANIE KHALFA	Chargé de recherche HDR, Aix-Marseille Université	Membre du jury
M. JACQUES DAYAN	Professeur des universités, Université Caen Normandie	Président du jury
M. FRANCIS EUSTACHE	Professeur des universités, Université Caen Normandie	Directeur de thèse
M. PIERRE GAGNEPAIN	Chargé de recherche Inserm HDR, Université Caen Normandie	Co-directeur de thèse

Thèse dirigée par FRANCIS EUSTACHE et PIERRE GAGNEPAIN, Neuropsychologie et Imagerie de la Mémoire Humaine

UNIVERSITÉ CAEN NORMANDIE

ÉCOLE DOCTORALE 556 HSRT

Normandie Université

Neuropsychologie & Imagerie de la Mémoire Humaine 1077

Remerciements

Je tiens à remercier la Dr Peggy Seriès et le Pr Fabien Vinckier pour avoir accepté d'être rapporteurs de ce travail de thèse. Je remercie également le Dr Khalfa et Pr Dayan d'avoir accepté d'examiner ce travail.

Je tiens à exprimer toute ma gratitude envers mon directeur de thèse, le Pr Francis Eustache, pour m'avoir accueilli au sein de l'Unité de recherche. Un grand remerciement pour m'avoir donné la possibilité de travailler sur un projet de recherche extraordinaire et « hors normes ».

Je tiens ainsi à remercier mon co-directeur de thèse, le Dr Pierre Gagnepain, pour avoir encadré ma thèse pendant ces quatre dernières années, pour son aide démesuré dans mon travail, pour tout le temps consacré à mes articles. Je souhaite aussi le remercier pour m'avoir initié au code, d'avoir dirigé des analyses parfois très compliquées, et pour les nombreuses et très stimulantes discussions scientifiques sur les aspects à la fois théoriques, cliniques et méthodologiques de ma thèse.

Je suis également très reconnaissant envers la Région Normandie, Normandie Université et l'INSERM pour avoir financé mes quatre ans de thèse, rendant ainsi possible ce travail.

Un sincère remerciement à toutes les personnes qui ont également permis de rendre possible la réalisation et le bon déroulement du protocole REMEMBER : Dr Denis Peschanski, Pr Jacques Dayan, Mme Carine Klein-Peschanski, Mme Florence Fraisse, Thomas Vallée, Dr Carine Malle, Dr Vincent de La Sayette, Dr Fausto Viader et Dr Emanuelle Duprey.

Un grand merci également aux post-docs, les doctorants et les neuropsychologues impliqués dans REMEMBER : Alison, Charlotte, David, Anaïs, Renaud, Benjamin, Céline, Lucie, Clarisse, Julie, Camille, Mélanie, Nelly et Gregory.

Je souhaite aussi remercier Florence Fraisse et tout le personnel administratif de l'Unité NIMH de m'avoir guidé dans les méandres obscurs de la bureaucratie lors de mon arrivée en France et durant toute ma thèse. Un grand remerciement aussi à tous mes collègues du PFRS et de Cyceron.

Je remercie Nicolas, Rémi, Prany, David, Joy, Francesca, Amalia, Alexis, Paul, Greg, Siya, Sanya, et tous-tes les autres ami(e)s qui m'ont accueilli et me font sentir chez moi dès mon arrivée à Caen.

Un ringraziamento speciale alla mia famiglia, mamma, papà e Erica, per essere sempre un sostegno infallibile e certo. Grazie ai nonni, agli zii e ai cugini. Grazie soprattutto alla nonna Giovanna: grazie per tutti i tuoi insegnamenti e per avermi invogliato sempre di più allo studio e alla conoscenza.

Pour finir, je remercie Cécile, qui partage ma vie. Merci de d'avoir supporté dans les moments de stress et de folie, de rentrer dans le bureau en dansant quand à le soir je travaillais encore. Merci pour ton soutien infini dans les moments les plus durs de ma thèse. Avec toi, et grâce à toi, chaque moment est heureux.

INDEX

Abbreviations' list

ACC: Anterior cingulate cortex

aMFG: Anterior middle frontal gyrus

BA: Brodmann area

BF: Bayes factor

BLA: Basolateral amygdala

BMA: Bayesian model averaging

BMS: Bayesian model selection

BNA: Brainnetome atlas

BOLD: Blood-oxygen-level-dependent

BOR: Bayesian Omnibus Risk

CA: Cornu Ammonis

cHIP: Caudal hippocampus

CI: Confidence interval

CS: Contextual stimulus

DCM: Dynamic causal modelling

DG: Dentate gyrus

dlPFC: Dorsolateral prefrontal cortex

DSM: Diagnostic and statistical manual of mental disorders

EEG: Electroencephalography

fMRI: Functional magnetic resonance imaging

GABA: γ-aminobutyric acid

GLM: General linear model

HGF: Hierarchical Gaussian filter

HPA: Hypothalamic–pituitary–adrenal

HS: Hippocampal subfields

IB: Imbalance angle

KF: Kalman filter

LTP: Long-term potentiation

MCMC: Markov Chain Monte Carlo

MEG: Magnetoencephalography

MFG: Middle frontal gyrus

MNI: Montreal Neurological Institute

NMDA: N-Methyl-D-aspartate

NT items: No-Think items

PC: Precuneus

PCL: Posttraumatic disorder check-list for DSM-5

PE: Prediction error

PFC: Prefrontal cortex

pMFG: Posterior middle frontal gyrus

PPI: Psycho-physiological interaction

pre-SMA: Pre-supplementary motor area

PTSD: Post-traumatic stress disorder

PXP: Protected exceedence probability

RF: Resultant force

RFX-BMS: random-effects Bayesian model selection

rHIP: Rostral hippocampus

RIF: Retrieval-induced forggetting

ROI: Region of interest

ROPE: Region of practical equivalence

RW: Rescorla-Wagner

SCID: Structured clinical interview for DSM-5

SIF: Suppression-induced forgetting

T1: Time 1

T2: Time 2

T3: Time 3

TH items: Think items

TNT: Think/No-Think

US: Unconditioned stimulus

vlPFC: Ventrolateral prefrontal cortex

WB: White Bear

wHIP: Whole hippocampus

XP: Exceedence probability

INTRODUCTION

1. Adaptive forgetting

The past shapes the present and the future. Memory traces of previous encounters with world constitute the grounds for the interpretation of the reality. When confronted with new situations, human beings have the adaptive ability to integrate past and present knowledge to produce a unified schema of the environment, useful to predict the future encounters with similar situations.

Philosophers have been debating the functioning of memory for millennia. Starting from Plato's definition of memory as wax tablet, in which perceptions and thoughts are impressed (Plato 369AD), the perceptual and emotional aspects of memories have been extensively developed in modern philosophy and science. A worthy example from literature is the *madeleine* of Marcel Proust, in his masterpiece "*À la recherche du temps perdu"* (Proust 1919). The French author masterfully describes reminiscences: by eating a biscuit and drinking a cup of tea, he involuntarily retrieves childhood's memories seemingly lost. Such memories were composed of both strong perceptual and emotional contents. A simple taste perception caused a cascade of visual reminiscences and incontrollable emotions. This early insight from literature lays the foundations for some major questions of contemporary research on memory: how do we store memories? Why does a certain external cue activate a specific memory, within a large spectrum of other possible memories? Why do we forget?

1.1. How memories are stored

At the beginning of 20th century, Richard Semon introduced the term "engram" to describe the neural substrates containing persisting experiential representations stored as a structure that can be retrieved, reused and updated (Semon 1921). Since these early intuitions, decades of research have been focusing on a "quest for the engram", investigating the neural and physiological substrates of memory storage and reactivation. In 1949, the physiologist Donald O. Hebb proposed that the concerted activation of different neuronal populations generated by perceptual experience facilitates its storage (Hebb 1949). His popular statement "*neurons that fire together wire together*" opened boulevards in memory research, providing

evidence for the long-term potentiation (LTP) a few decades later. LTP is a cascade of physiological mechanisms, mostly depending on the postsynaptic N-Methyl-D-aspartate (NMDA) receptors, leading to different hippocampal neural populations to synchronize their firing activity during learning, strengthening their synapses and, consequently, facilitating the memory storage (Malenka and Nicoll 1999). There is today a wide consensus on defining LTP as the physiological substrate of the memory engram formation.

Memory traces can last a few seconds as well as a whole life. Typically, a distinction is made between working memory, short-term memory and long-term memory. Working memory refers to the online attention-related mechanisms enabling retaining salient information for a few seconds when processing external stimuli. The distinction between short-term and long-term memory essentially relies on their *duration* and their *capacity*, defined as the time the memory traces can last and the amount of information that they can contain, respectively (Cowan 2008). While short-term memories can last a few seconds to a few minutes and have strong capacity limits, long-term memories can last days to decades and the amount of information they can contain is potentially unlimited. Whether the link between these two types of memory is sequential (i.e., short-term memories become long-term memories) or these two types of memory reflect two independent and parallel processes is still debated (McGaugh 2000).

Three different key memory processes should be distinguished in order to understand how short-term memories could eventually be converted into long-term memories: consolidation, retrieval and reconsolidation (see Figure 1). **Memory consolidation** is the process enabling memories to be stored. From a neurophysiological point of view, consolidation has long been identified with the Hebbian strengthening of synaptic connections within a network of neuronal populations. Popular models propose a brain hierarchy in different phases of the memory consolidation. In a review of the literature, Meeter and Murre (2004) proposed an interesting model of memory consolidation. Accordingly, information is initially held in working memory structures situated in the prefrontal cortex for about one minute. If the memory has to be retained for longer, a cascade of cellular consolidation mechanisms starts in the hippocampus, allowing the memory to be stabilized and stored in this subcortical structure. Consequent long-term memory consolidation depends on the strengthening of the memory trace in the neocortex. This model has found strong evidence in amnesia studies showing that lesions in the hippocampus cause anterograde, but not

retrograde, memory loss, suggesting different time-dependent roles of the hippocampus and the neocortex in the stabilization of memories (Meeter and Murre 2004).

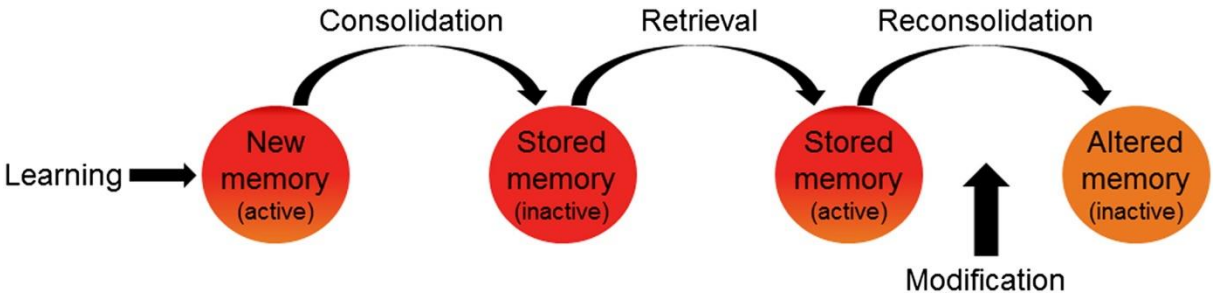


Figure 1. Memory consolidation, retrieval and reconsolidation. Adapted from Schwabe, Nader, and Pruessner (2014).

Memory consolidation is tightly related to the concepts of **memory retrieval and reconsolidation**. Consolidated memories have long been considered as fixed and unlikely to be lost in normal conditions. However, recent evidence suggests that consolidated memories can be modified. Accordingly, consolidated memories are stored in an inactive state, preventing any modification of their contents. However, when memories are retrieved, they shift into an active state, vulnerable to modifications: like the past shapes the present, the present reshapes the past. At the neurophysiological level, memory retrieval destabilizes the neuronal pathways that have potentiated during consolidation of the memory trace (Schwabe et al. 2014). This destabilization let the memory vulnerable to be restructured, updated and altered before being stored again, a process known as reconsolidation (see Figure 1).

In a stimulating study on memory reconsolidation, Hupbach et al. (2007) demonstrated that the exposition to reminders plays a key role to the modifications of memories. In their experiments, healthy participants learned lists of objects at day 1 and at day 2. Crucially, at day 2, before learning the second list, a subgroup was exposed to reminders of the day one's list while another subgroup was not. When, at day 3, participants were asked to recall the day 1 objects, the subgroup exposed to the reminders incorrectly mixed the two days' object lists, while the other subgroup did not. These results showed that memories became labile and sensitive to changes when elicited by reminders, producing strong evidence for reconsolidation mechanisms.

1.2. Forgetting is fundamental for remembering

Decades of memory research have been focusing on the mechanisms underlying how we remember. However, remembering is not possible without forgetting. Forgetting has long been considered as a passive, often undesirable, mechanism in opposition with remembering. Since the dawn of the neuropsychology, forgetting has been seen as a failure of memory, as manifest in some neurological conditions, such as amnesia and dementia.

Only in the last twenty years a growing body of research carried evidence that forgetting is not only passive, but it is rather a fundamental mechanism for memory, and it is constantly at work in daily life (Gravitz 2019a). Human beings are constantly exposed to a multitude of external stimuli and retaining such amount of information would be extremely costly. Furthermore, not all memories are welcome: some memories can be extremely distressing and their recall unwanted. Early insights in philosophy and psychoanalysis advocated the importance of forgetting, considering it as an active and motivated process, rather than a passive memory deficiency. Nietzsche firstly recognized the role of forgetting in the prevention of negative emotions associated with unwanted memories (Nietzsche 1886):

"Blessed are the forgetful; for they get the better even of their blunders"

Friedrich Nietzsche,

Beyond Good and Evil: Prelude to a Philosophy of the Future

Nowadays, numerous beneficial sides of forgetting are ascertained. Besides allowing to evade the emotional consequences of the past, forgetting also allows removing inconsistent information in order to build a coherent internal model of the world, as well as avoiding distractions, minimizing the competition of contrasting and redundant memory details, improving creative solutions to problems and motivating the reconnection with the past (see Fawcett and Hulbert 2020 for an exhaustive review). For long time memory researchers have been focusing on how memories are formed and stored, but this approach may depict only one toss of the coin: to fully understand how we remember, we should before understand how we forget (Gravitz 2019b). There are two main classical theories of forgetting, historically

viewed as competitive: memory decay and memory interference. We will also focus on a third, recent theory modelling forgetting as an active mechanism.

1.3. Memory decay theory

Memory decay theory proposes that memory traces undergo natural deterioration over time. In other words, we forget because, due to storage capacity limits, memory traces disintegrate with the disuse in time. The first cornerstone for the decay theorization was the description of a time window in which memories are more vulnerable to be lost, in the early phase of acquisition, by Ribot (1882). Studying patients with amnesia, the French psychologist proposed that some time is needed before the memory can be stabilized through the consolidation process and that, during such time, memory traces are vulnerable to disintegrate, hypothetically explaining why in amnesia recent memories are more likely to be lost than more ancient memories. Similarly to Ribot, Thorndike (1913) formulated his law of use and disuse, stating that when an association is not made between a stimulus and a response under a determinate amount of time, such connection is less likely to be established later in time.

In an experimental study conducted in 1958, John Brown firstly suggested that the time-related decay of memory traces can occur not only for long-time memories, but also for short-term memories (Brown 1958). In Brown's paradigm, extensively replicated afterward, participants were exposed to lists of letters of different length and after a delay of three-to-30 seconds they were asked to recall the order of such pairs. Across decades, the results of studies using this paradigm showed that a few seconds were sufficient to observe the decay of memory traces. According to the decay theory, forgetting is a failure of memory consolidation (Lewandowsy 2010). The decay theory has rapidly became quite unpopular, with stronger evidence in support of the "competitive" interference theory (see below).

Only in recent years, along with new discoveries in molecular neuroscience and substantial technical advance in neurobiology, this theory is regaining popularity. Novel derivations of the decay theory go beyond the conception of decay as a passive mechanism, as originally formulated, comparable to the radioactive decay, moving towards memory decay as an active process. It has been recently proposed that memory decay is a well-organized neuronal process conceived to remove the numerous irrelevant memory traces stored during

the day (Hardt, Nader, and Nadel 2013). Accordingly, such unnecessary amount of hippocampal-dependent daily memory traces undergoes selective deletion during sleep, a temporal window in which the brain is not engaged in learning new information. Novel decay theories propose that the loss of long-term consolidated memories, for which the neocortex plays a central role (see above), target the neurobiological substrates of the spatial-contextual components of memories, precluding the retrieval of the memory content stored in the neocortex.

1.4. Memory interference theory

While the decay theory assumes that memory traces dissolve by nature, the interference theory proposes that forgetting is due to competition between different memory traces. This phenomenon was first described more than a century ago by Müller and Pilzecker (1900). These authors have found that recalling a memory associated with a cue was less likely when the same cue was afterward associated with another competitive memory. This first account of interference between different memories was originally thought to happen because the new memory trace interrupted the consolidation of the old memory. Although this consolidation-dependent explanation of memory interference was quickly abandoned, decades of research focused on this phenomenon to understand forgetting, in contraposition with the memory decay theorists (for a historical review, see Wixted, 2004).

Two different, yet complementary, types of memory interference should be disambiguated. In typical interference task, subjects learn a list of associations between cues *A* and responses *B*, and then a second list containing the same cues *A* associated with the responses *C*. The following interference phenomena can be observed:

- Retroactive interference: being exposed to the second list *A-C* impairs the recall of the *A-B* list's associations;
- Proactive interference: being exposed to the first list *A-B* impairs the recall of the *A-C* list's associations.

These two interferences, both contributing to forgetting, can be view as two sides of the same mechanism. In the interference theory, there is no role for memory consolidation in

forgetting (Lewandowsy 2010). According to this theory, memory decay because new or old memories interfere.

Given two competitive memories associated with the same cue, one of the two associations is stronger and naturally prevails over the other. However, in the real life, sometimes the weaker memory is more appropriate in a given contest and the stronger memory response has to be inhibited (Anderson 2003). Despite the early insights on competitive memories as a factor of forgetting, the classic interference theory left an unaddressed key question: how memories can be actively selected and inhibited? Regardless of the noticeable differences between decay and interference theories, both originally assumed forgetting being not an active mechanism, but rather an incidental process due to resources limitations. Originated from the evolutions of the interference theory, the more topical concept of active forgetting offers strong evidence of forgetting as a voluntary process, possible to be intentionally directed (Anderson and Hulbert 2021).

1.5. Active forgetting

Certain memories, at times, can be unwelcome. As in the case of Proust's madeleine, memories can have strong and vivid emotional contents. Memories can have negative, stressful and undesirable connotations, making them non-adaptive in some contexts. Furthermore, sometimes the situation requires the recall of a specific memory, or to focus in a task without recalling memories, and recalling alternative memories could be inappropriate and distracting. As uncontrollable and stressful intrusive memories are key features of some psychiatric disorders (Reynolds and Brewin 1999), understanding how memories can be voluntarily directed and forgotten is crucial.

In the everyday life, different memories can be activated by a single environmental cue at which they have previously been associated. However, only one of these possible alternative memories may be appropriate in the ingoing contextual demands. This range of possibilities is an important, extensively studied, characteristic of goal-directed behaviour: living beings continuously select their behavioural responses, more or less implicitly, in order to optimize the successfulness of their interaction with the environment, maximizing their chances to survive. **Control mechanisms** are fundamental for directed behaviour, and they are characterized by two main complementary aspects: the selection of the most appropriate

behavioural response to reach the current goals and the inhibition of the competitive behavioural alternatives. The control of goal-directed motor and behavioural responses is strongly dependent on the prefrontal cortex (Mostofsky and Simmonds 2008).

It has been proposed that the regulation of the internal cognitive states, including memory retrieval, shares selection and inhibition processes with the directed behaviour mechanisms (Anderson 2003). Accordingly, when confronted with a reminder cue potentially eliciting multiple memories, active control mechanisms ensure the selection of the most pertinent memory in line with the contextual demands and the parallel inhibition of the other competitive memories. Two forms of active forgetting have been investigated in the last three decades: retrieval-induced and suppression-induced forgetting.

The core idea underlying **retrieval-induced forgetting (RIF)** is that remembering can cause forgetting. Originated from the interference theory, the idea of RIF goes beyond the view of forgetting as a passive mechanism. While the interference theory postulated that acquiring new memories impair older memories, and vice-versa, because of limits in storage capacities, RIF implies that forgetting is the consequence of the active inhibition of competitive memory traces (Anderson 2003). In a classic RIF paradigm, namely retrieval-practice, participants are instructed to learn a list of category-exemplar associations (e.g., fruit-orange, fruit-apple) and then they practice the retrieval of half of the associations' list. After a 20-minutes interval, participants perform a recall test for all the associations. Early, extensively replicated, studies using this paradigm have shown that unpractised associations were harder to recall, suggesting that retrieving a memory weakens the competitive memories associated with the same cue, making them less accessible to future recall (Anderson, Bjork, and Bjork 1994).

However, one may argue that this retrieval-related forgetting could be due to the fact that retrieving a memory makes it more accessible, via the potentiation of the synapses linking the neural substrates encoding its trace. In this point of view, non-practiced items would not be forgotten, but, rather, the practiced items would gain strength by practice, resulting in an imbalanced memory competition. Contradicting this objection, evidence show that strengthening a memory does not affect the competitive memories (Anderson and Hulbert 2021), and no correlation has been found between RIF and the strengthening of practiced memories (Hulbert, Shivde, and Anderson 2012). The foremost evidence in favour of RIF is its independence from the cues: RIF of competing memories is persistent when testing the

unpractised memories for other unrelated categories, thereby demonstrating that the competitive memory traces are weaken and less accessible to recall, in spite of the associated cue and the strengthening of the practiced memories. Importantly, RIF is a general forgetting mechanism and it has been found in various memory domains, ranging from visual objects to autobiographical memories (for an exhaustive review, see Anderson and Hulbert, 2021).

Also the direct of a memory trace can cause forgetting. While SIF provides evidence that selecting specific memories impairs the competitive memories, **suppression-induced forgetting (SIF)** implies directly inhibiting unwanted memories. When unwanted memories intrude consciousness, people often attempt to stop such unwelcome retrievals. The main idea underlying SIF is that directly suppressing memories via inhibitory control provokes their weakening, making later recall less likely (Anderson and Hulbert 2021). The mechanisms of such form of forgetting have been extensively studied in the last two decades, mainly through the Think/No-think Task. Given the central role of SIF in the current work, a detailed account is needed. In the next chapter, the focus will be laid on the behavioural findings and the neural basis of memory suppression.

2. Memory suppression

Evidence shows nowadays that human beings are gifted with the ability to actively direct forgetting. Differently from the theorization of forgetting due to decay, interference, or retrieval, studies on memory suppression focus on the ability to voluntary suppress undesired memories via inhibitory control. In the current chapter attention will be paid on the principal memory suppression experimental paradigm, the Think/No-think task (TNT), and on experimental research investigating the behavioural and neuronal basis of memory suppression in humans.

2.1. The Think/No-think task

Developed by Anderson and Green (2001) and extensively replicated in the last two decades, the TNT allows evaluating the distinct, yet complementary, inhibitory control mechanisms underlying both preventing unwanted memories from intruding consciousness and purging away undesired memories. The TNT follows the same principles of a typical Go/No-Go task, an inhibitory control motor task. Besides relevant experiment-related specificities, the TNT paradigm is commonly structured into three different phases:

- **Learning phase:** participants are asked to learn a list of cue-target pair associations. Pairs can be words, pictures, or name-object associations (for example, see **Figure 2**, on the left). These association pairs are repeatedly presented and learning is tested by presenting the name and asking the participants to recall the associated object, until a learning criterion is reached, usually 90% of correct associations.

- **TNT phase:** after the learning phase, when associations are mastered by the participants, the real task begins. In this phase, only the cues are presented, and the participants are asked either to recall the paired item (Think items, TH) or to focus on the cue and prevent the associated item from entering consciousness (No-Think

items, NT, see Figure 2, on the middle). Importantly, if the object intrudes consciousness during the NT condition, participants are asked to purge away it away. In some variants of the TNT task, participants report whether they experienced or not memory intrusions at the end of each trial (Levy and Anderson 2012a).

- **Test phase:** Following the TNT, either a recall test or a priming test is performed to assess the accessibility of the targets associated with the cues, allowing measuring the effect of suppression on forgetting by comparing the availability of the memory trace (see Figure 2, on the right). This phase allows measuring the availability of the memory trace for items that have been repeatedly retrieved (TH items), items that have been repeatedly suppressed (NT items) and items that were studied but was not presented in the TNT phase (baseline items).

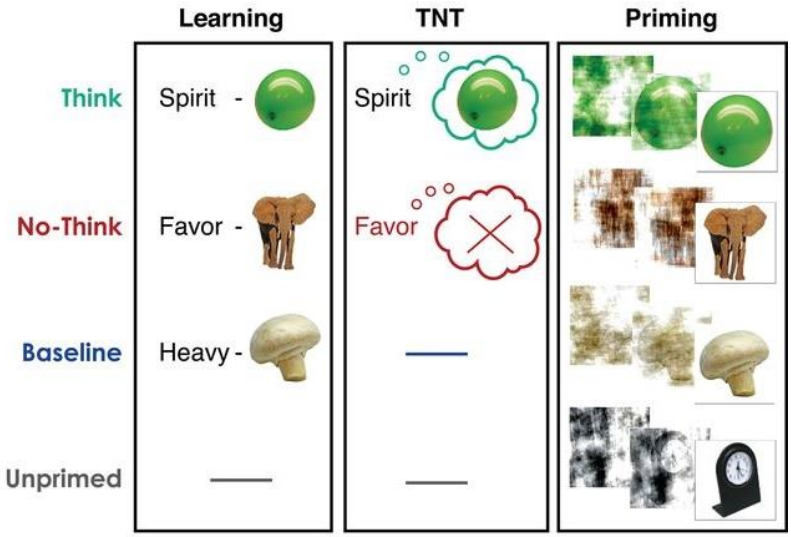


Figure 2. The Think/No-think task. After learning associations between names and objects (learning phase), only names are presented and the participants are asked either to think or not to think to the associated object (TNT phase). A recall test or a priming test is then administered to evaluate the forgetting effect of memory suppression on the accessibility of suppressed memory traces. Adapted from Gagnepain, Henson, and Anderson (2014).

The TNT task elegantly accounts for memory suppression, presenting substantial specificities and advantages compared to other analogous paradigms. This task allows at the same time the active prevention of intrusive memories and the active purging of unwanted memories intruding consciousness. The main findings, discussed in detail below, suggest that suppressed memories are less accessible during later recall, demonstrating SIF. Since determining whether a person truly prevent intrusive memories cannot be objectively

observed, the act of actively attempting to stop an intrusive retrieval, as required by the TNT, is crucial for probing the neurocognitive inhibitory mechanisms underlying memory suppression (Anderson and Huddleston 2012).

Other paradigms have been implemented in the past decades to study how thoughts are avoided and suppressed, presenting substantial differences with the TNT. An early classic example of thoughts suppression task is the "white bear" (WB) paradigm, in which people are asked to do not think to a white bear for five minute and to ring a bell if they think about it (Wegner 1994). Obviously, this instruction elicits the image of a white bear into awareness, and participants should suppress it. Paradoxically, empirical evidence has shown that attempting to suppress the WB results in a rebound of this thought, which becomes even more accessible. This effect often brought researchers to the conclusion that suppressing thoughts is paradoxically counterproductive (Wenzlaff and Wegner 2000). However, these conclusions can be biased by the fact that the WB paradigm implicitly requires the maintaining of the to-be-suppressed trace into awareness, creating a conflict between the instruction and the task (Anderson and Huddleston 2012).

The TNT task consistently goes beyond this strong limitation, as there is no direct reference to the to-be-suppressed object. As the object to suppress is eventually reminded by an unrelated cue, this task allows isolating genuine memory suppression mechanisms. Memory suppression should also be differentiated from cognitive avoidance. Cognitive avoidance imply the active circumventing of any reminders of unwanted thoughts and memories, leaving their traces unaltered, and it is often associated with adverse psychological outcomes and psychiatric conditions. On the contrary, controlling memories involves the direct exposition to reminder cues and the consequent cognitive active strategies deployed to suppress the associated unwanted memories or thoughts can be the cause of the degradation of their traces (Engen and Anderson 2018).

2.2. Behavioural findings

Numerous studies on memory suppression have been conducted through the TNT paradigm, pointing out a crucial result: suppressed memories are less likely to be recalled. In the first TNT study, Anderson and Green (2001) have found that suppressed items were harder to recall when compared to TH and baseline conditions' items, showing for the first

time a direct effect of suppressing memories on their later recall. In this first seminal study, the items were constituted of noun-noun association pairs and when reminder names were presented participants were asked to either recall or suppress the associated names. In order to further isolate the effect of inhibition, to test the hypothesis that suppressing an unwanted memory impairs the memory itself, rather than the strength of its association with the reminder cue, the researchers implemented an independent probe test. During this test, the target names were elicited through novel semantic cues (for example, for the target word "roach", subjects received 'insect r___' and they were asked to recall the previously learned word fitting the cue). Crucially, the memories suppressed during the TNT phase were harder to be recalled also when elicited by an external cue. These results have shown that the effect of memory suppression in forgetting is independent from the reminder cue. Accordingly, memory suppression does not simply affect the association between a reminder cue and a memory trace. Rather, memory suppression directly targets the degradation of the suppressed memory trace, making it less accessible to future recall. These early results inspired two decades of studies on memory suppression.

Ten years after the first evidence of memory suppression, a review of 32 published TNT studies involving a total of 2,174 participants confirmed that NT memories were significantly less often recalled than TH and baseline items, when elicited by both same-probe (i.e., a simple recall test using the previously learned associated cues) and independent-probe cues (Anderson and Huddleston 2012). Across experimental studies, suppressed memories are harder to recall compared to baseline memories – a phenomenon known as negative control effect – and TH items are more likely to be recalled than baseline memories – a phenomenon known as positive control effect. Interestingly, it has been found a correlation between the overall frequency of intrusive memories and later forgetting, suggesting that forgetting was more likely in individuals who were more capable to purge intrusive memories from consciousness (Levy and Anderson 2012a).

Two strategies can be deployed to control memories when confronted to reminders: direct suppression and thought substitution. The difference between these two strategies lies in the fact that, while in direct suppression people are asked to focus on the cue and avoid distractors, in thought substitution paradigms the participants are asked to prevent the unwanted memory by actively redirecting the focus of attention on their own distractor thoughts. Both strategies lead to increased forgetting of the suppressed memories (Anderson and Huddleston 2012). Beyond the questioned strategy, individuals' ability to suppress

memories during the TNT task predicts their performances at later recall test, showing a linear positive relationship between the ability to suppress memory and the amount of SIF (Levy and Anderson 2012a).

In some variants of the TNT task, a perceptual identification test is performed instead of the classic recall test. Objects associated with names belonging to NT, TH and baseline conditions, as well as new objects, are presented with gradually reducing visual noise and participants are asked to recognize the objects as fast as they can. Being exposed to an object generally facilitates its recognition later on. Experimental results have shown that suppressed objects were harder to perceptually identify compared to TH and baseline items, as shown by longer reaction times. Thus, suppressing memories might disrupt the sensory component of memory traces, thereby reducing their influence on later perception (Gagnepain et al. 2014). Furthermore, suppressing emotionally unpleasant memories also reduce their later affective valence. A TNT using emotional materials has found that participants who were effective in preventing negative intrusive memories were likely to assign lower negative valence to the suppressed unpleasant stimuli after the task (Gagnepain, Hulbert, and Anderson 2017). These results suggest that suppressing unwanted memories could affect both memory traces and the associated emotional traces. The most plausible explanation for SIF is that retrieval stopping is accomplished via specific inhibitory mechanisms targeting the disruption of the memory traces. Many neuroimaging studies contributed in identifying and well-characterizing a brain network responsible for memory suppression, strongly supporting this hypothesis.

2.3. Neural bases

Similarly to other forms of control, memory suppression engages a brain network mostly involving the prefrontal cortex (PFC). When engaged in a memory suppression task, such control regions suppress the activity of a large memory network, stopping the retrieval of unwanted memories. Two core networks interplay in an inhibitory fashion: control network and memory network.

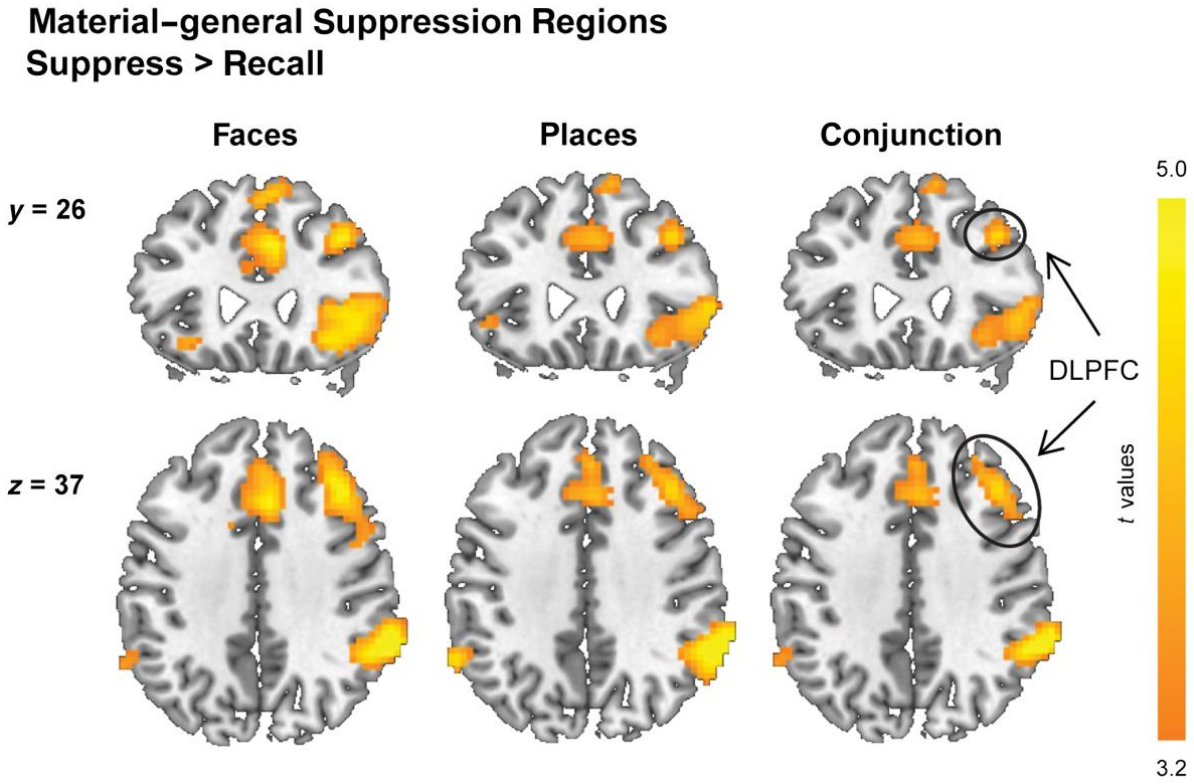
Control regions

The **PFC** plays a central role in executive functions and controlling behaviour, cognitive states and motor responses, as well as to select and direct the most appropriate behaviour in a given context. Such high-order functions often require the brain to inhibit the dominant response to achieve goals. For example, imagine a mind-wandering person walking towards a room whose door is seemingly open but, suddenly, just before crashing on it, this person realizes that there is glass door. This situation necessitates an immediate stopping of the motor action, overriding a reflexive, preponderant ingoing motor action. The mechanisms underlying motor stopping have largely been studied through the go/no-go task, the equivalent of the TNT task for the motor domain, were subjects are instructed to either go ahead or to inhibit a motor response. Evidence has shown that the activation of the PFC is responsible for motor inhibition (Levy and Wagner 2011).

Likewise stopping walking to avoid crushing into a glass door, preventing and stopping an unwanted memory intruding consciousness requires prefrontal inhibitory control. Neuroimaging studies have often investigated the brain mechanisms of memory suppression by contrasting NT and TH conditions of the TNT task. Greater brain activations during the NT compared to the TH condition unveil the neural system responsible for memory suppression, above and beyond the brain processes involved in memory retrieval, which, on the contrary, are well-identified by investigating the regions more activated during TH than NT (Anderson, Bunce, and Barbas 2016). In the first TNT functional magnetic resonance imaging (fMRI) activation study, Anderson et al. (2004) have found increased activations in a neural system including bilateral dorsolateral and ventrolateral PFC (dlPFC and vlPFC, respectively) as well as the anterior cingulate cortex (ACC) and the pre-supplementary motor area (pre-SMA). This early investigation contributed formalizing memory suppression as an active process involving regions well-known to contribute to executive functions, such as motor stopping.

Many studies replicated similar results, often with an important additional finding: the brain memory control system is strongly **right-lateralized**. Several fMRI studies have reported the broadest activations in the right dlPFC, extending in both the anterior and posterior portions of the middle frontal gyrus (MFG), suggesting a cluster formed by the Brodmann areas (BA) 9, 46 and 10 as the key locus for memory suppression (Depue, Curran, and Banich 2007). Other frequently reported activated regions include a large cluster in the

midline frontal cortex ACC, the pre-SMA and right parietal regions (see Figure 3 and Anderson, Bunce, and Barbas (2016) for an exhaustive review). Crucially, the activation of this brain control network has been observed across various types of stimuli and task materials. For example, in a study by Benoit et al. (2015) participants had to suppress visual objects' memories depicting faces or places. The activation of the control network in these two different conditions almost completely overlapped, demonstrating that the neural system engaged to suppress memory is highly general and independent from the nature of the memories to suppress (see **Figure 3**). Indeed, a common control network inhibits different



target brain regions according to the different nature of the stimuli (see below).

Figure 3. TNT brain activations. Increased fMRI activations during NT condition compared to TH condition at the TNT task. Suppressing memories activate a large right-lateralized fronto-parietal network. On the left, increased activations when visual memories to suppress are faces; on the middle, increased activations when visual memories to suppress are places; on the right, the overlapping of brain activations between "faces" and 'places" cues. Adapted from Benoit et al. (2015).

There are several reasons why the **right dlPFC** has been considered the key region of memory suppression. First, although increased activity in other frontal regions have been reported when contrasting NT with TH activations, such activity might be not specific to memory suppression. Two distinct processes can occur during the NT condition: on the one hand, some processes could target the prevention of memory intrusions and, on the other

hand, other processes could target the suppression of intrusive memories that already entered consciousness (Benoit et al. 2015). Specifically, NT trials can be divided into two categories: intrusive and non-intrusive trials. While the left PFC is activated only during non-intrusive trials (i.e., when the system is engaged in preventing intrusions), the right dlPFC is selectively more activated when the brain faces intrusions to be purged away (Benoit et al. 2015). Second, the vlPFC and the ACC, but not the dlPFC, are activated during thoughts substitution, another form of prevention of intrusive memories, additionally suggesting the specificity of the dlPFC in memory suppression (Benoit and Anderson 2012). Third, the strength of the right dlPFC, activation during the NT condition is a significant predictor of the degree of SIF (Anderson et al. 2004; Depue et al. 2007). Fourth, as evidence suggested that the dlPFC is highly implied in other forms of control, including motor stopping, emotional regulation and cognitive control, this region is proposed to be a supramodal inhibitory region targeting different task-dependent activation sites (Depue et al. 2016). Taken together, these results suggest that a large right frontoparietal is engaged when people suppress unwanted memories, with a specific key role of dlPFC.

Suppressed regions

While increased prefrontal activity observed in literature during memory control closely overlaps the motor control regions, the two sensitively differ on the regions whose activity is reduced during control. While in motor control tasks reduced activity in the primary motor cortex has been observed, not surprisingly, **medial temporal lobe** regions, fundamental for memory retrieval, show reduced activity during memory suppression (Anderson et al. 2016). Given its central role in memory retrieval, the **hippocampus** is the elected target whose activity needs to be suppressed when retrieval has to be stopped. Consistently, fMRI TNT studies have shown substantial reductions in the bilateral hippocampal activity during memory suppression compared to memory retrieval (Anderson et al. 2004; Benoit et al. 2015; Depue et al. 2007). However, these results could be interpreted as a simple increasing in hippocampal activity during memory retrieval, rather than an inhibition during suppression. Contradicting this objection, the hippocampus showed reduced activity also when memory suppression is compared to a cross fixation baseline condition (Depue et al. 2007). Furthermore, a correlation has been found between the increased activity of dlPFC and the

decrease of hippocampal activity, suggesting, but not proving, a potential inhibitory interplay between these two regions (Benoit et al. 2015; Benoit and Anderson 2012).

A functional and anatomic distinction can be done between the anterior and the posterior hippocampus. The anterior hippocampus is specialized in episodic memory encoding and posterior hippocampus in memory retrieval, and it has been proposed that, for its role in retrieval, the posterior hippocampus could be the foremost target to inhibit during memory suppression (Anderson et al. 2016). Regardless its subparts, a specific reduction in hippocampal activity has been found in intrusive trials when compared to non-intrusive trials, showing that the downregulation of this region is important when unwanted memories intrude consciousness and need to be purged; moreover, this hippocampal deactivation correlate with later SIF (Levy and Anderson 2012a).

Other regions have been shown to reduce their activity during memory suppression, such as bilateral temporal lobes' perirhinal and enthorinal areas, the posterior cingulate cortex, and parietal regions like the retrosplenial cortex (see Anderson, Bunce, and Barbas (2016) for a review). Crucially, reductions in some brain areas' activity during memory suppression compared to retrieval depend on the nature of the stimuli used in the TNT task. When the memories to suppress are visual objects, visual cortex undergoes activity reduction (Gagnepain et al. 2017); by contrast, this decrease was not observed for verbal memories (Depue et al. 2007). Furthermore, reductions in the activity of the amygdala have been observed when the memories to control have negative emotional connotations (Gagnepain et al. 2017). Similarly, reduced parahippocampal activity has been specifically observed when stimuli are places (Benoit et al. 2015), and face recognition area in the fusiform gyrus deactivation has been specifically observed when the stimuli to suppress are human faces (Benoit et al. 2015).

Taken together, these results indicate that a prefrontal control system, notably the right dlPFC, exerts a form of inhibitory control over memory areas, targeting the multiple brain areas involved in the retrieval of the different features of the memory to suppress (e.g., visual, emotional, etc.). However, although this hypothesis is plausible, fMRI activation studies do not allow inferences on how different brain regions communicate during memory suppression. In order to validate this inhibitory control hypothesis, fMRI studies have been focusing on the functional connectivity between control and memory regions.

Brain Connectivity

The brain is a complex organ whose high-level functions depend on the interplay of different specialized networks and multimodal associative areas. Each brain population presents two fundamental characteristics: segregation and integration. While segregation refers to the functional specialization of brain areas in processing some specific aspects of external or internal stimuli, integration refers to the fact that complex brain functions depend on the connections between different specialized areas (Friston 2011). Functional connectivity methods allow measuring the concerted activation of different brain regions. Different methods have been implemented, basing for example either on time-course correlations between different regions when the brain is engaged in a task, or in modelling the directed effective connectivity (for a complete review, see Friston, 2011).

Several research studies have examined the relationship between the control and the memory systems during memory suppression. TNT studies investigating the functional connectivity through the Psycho-Physiological interaction (PPI), a method based on the general linear model (GLM) assumptions, have generally found a negative relationship between the dlPFC and the hippocampus (Benoit and Anderson 2012; Gagnepain et al. 2017; Liu et al. 2016). In these studies, the increased activity of the dlPFC correlated with the decreased activity of the hippocampus. Although these results are fascinating and suggestive of a prefrontal inhibitory control over the hippocampus in memory suppression, due to their correlational nature, these connectivity methods do not allow inferences on the directionality of the relationship between control and memory systems.

Strong evidence on the directionality of the relationship between the dlPFC and the hippocampus arose from studies addressing this question with Dynamic Causal Modelling (DCM). This technique allows inferring changes in effective, directed connectivity by building and comparing different hypotheses-driven generative models of neural dynamics underlying fMRI activations (see Friston, Harrison, and Penny (2003) and the paragraph **Dynamic Causal Modelling** for a detailed description). Several DCM studies have demonstrated that memory suppression engages **dlPFC-orchestrated inhibitory control over the hippocampus**, as showed by negative coupling between these two regions (Benoit and Anderson 2012; Gagnepain et al. 2014, 2017). Crucially, after testing several alternative hypotheses, across different studies, evidence has shown that the direction of this negative coupling is top-down, with dlPFC directly inhibiting the activity of the hippocampus and

other regions responsible for memory retrieval. Depending on the nature of the to-be-suppressed memories, the dlPFC inhibitory control has been observed towards brain regions specifically associated with the memory cue characteristics. For example, a study by Gagnepain, Hulbert, and Anderson (2017) has shown that, in parallel with the hippocampus, inhibitory control also specifically targets the amygdala when the memory to suppress has a negative valence.

Inhibitory control increases during intrusive trials when compared to non-intrusive trials. Although top-down downregulation has been observed when participants prevent intrusive memories, greater inhibition occurs when unwanted memories enter consciousness (Levy and Anderson 2012a). Such increased downregulation of the hippocampus observed in intrusive trials could appear counterintuitive at first glance, but is a clear marker of the mechanisms deployed to purge away intrusions. Some studies have found that the greater hippocampal downregulation during memory suppression correlated with increased amount of forgetting (Benoit and Anderson 2012) and reduced perceptual identification of the suppressed memories (Gagnepain et al. 2014). However, when comparing intrusive and nonintrusive trials, Levy and Anderson (2012) have reported that only the hippocampal downregulation during intrusive trials was predictive of future forgetting. Thus, the higher demand of inhibitory control required by intrusive memories entering consciousness is associated with improved downregulation of the regions supporting memory retrieval and, consequently, higher degrees of SIF.

Although connectivity and DCM studies shed light on the fundamental role of prefrontal cortex inhibitory control over the hippocampus for SIF, the neurobiological underpinnings of such processes are still poorly understood. A recent study combining fMRI and spectroscopy has found that hippocampal concentration of γ-aminobutyric acid (**GABA**) predicted the strength of the top-down connectivity from the dlPFC to the hippocampus (Schmitz et al. 2017). The integrity of the fronto-hippocampal inhibitory control network may depend on the GABAergic interneuron targeting the hippocampal inhibition. An opened, debated research question concerns the anatomo-functional pathways enabling the brain control regions to inhibit the hippocampal activity. Recently, Anderson, Bunce, and Barbas (2016) identified two possible distinct, yet complementary, pathways. According to the authors, prefrontal cortex may initially modulate the activity of entorhinal cortex, which, in turn, would constitute a "gate" depriving the hippocampus of its inputs from neocortical regions and, thus, preventing memory retrieval. On a second stage, if the entorhinal gating

fails resulting in intrusive memories entering consciousness, a second mechanism could target directly the hippocampus. This reactive control would be reached via inhibitory thalamic projections, notably originating from the thalamic reuniens nucleus. This theoretical model would differentiate two types of memory control: a *proactive* entorhinal control aiming to prevent intrusions and a *reactive* thalamus-dependent control directly targeting the hippocampus.

Taken together, many studies on memory suppression have shown that human beings can voluntarily attempt to suppress unwanted memories via top-down inhibitory control of the hippocampus and other retrieval-related brain regions, orchestrated by the dlPFC. These discoveries have important implications for advancing our understanding of psychiatric disorders characterized by intrusive memories or images, potentially shedding a new light on their aetiology. The presence and persistence of incontrollable intrusive memories is the key feature of post-traumatic stress disorder (PTSD, Reynolds and Brewin 1999). The next chapter will address the fundamental clinical characteristics of PTSD, and the central role of intrusive memories and their control in this psychiatric disorder.

3. Post-traumatic stress disorder

3.1. A brief history of PTSD

Symptoms nowadays falling under the diagnostic category of PTSD has been described since millennia. More than 2000 years ago, classical thinkers such as Hippocrates, Herodotus and Lucretius already described mental symptoms due to the exposure to battleships. In 440 B.C., Herodotus wrote about a brave soldier surviving the battle of Marathon who developed blindness without suffering any injury during the battle. In 40 B.C., Lucretius described a soldier re-experiencing the violence of the battleship, mostly through terrible nightmares (for a review, see Crocq and Crocq, 2000).

Dawn on modern times, physical explanations have been quested to understand such trauma-related symptoms. During industrial revolution, the developing railway often produced serious accidents. In this context, a medical condition known as "railway spine" grouped a variety of organic symptoms in victims who apparently did not suffer any physical injury (Harrington 2003). While during French revolution and Napoleon's wars some of the nowadays recognized PTSD symptoms were named "*vent du boulet*", or wind of cannonball, during World War I such symptoms were known as "shell shock", suggesting that the exposure to the artillery's violence could cause a pattern of physical and psychological symptoms, depicting a nervous soldier developing depression, tremor, inability to do anything (Loughran 2012). The origin of these symptoms has been long-standing debated. While most clinicians offered materialistic explanations implying somatic damages, some pioneering clinicians proposed a non-somatic interpretation, pointing out at the psychological aspects of the trauma, paving the way of modern psychiatry (for a review, see Harrington, 2003).

Even if in the aftermath of the World War II, with the publication of the Diagnostic and Statistical manual (American Psychiatric Association 1952), for the first time acute exposure to stressors was recognized to be the origin of a veterans' disorder characterized by anxiety, re-experiencing and sensitivity to trauma reminders, the turning point on the recognition of a trauma-related psychiatric condition arrived in the 70s of the past century

(Andreasen 2010). The United States of America were engaged on a challenging conflict in Vietnam, and activists redirected the attention of the government on post-war psychiatric symptoms, which took the name of "post-Vietnam syndrome", and did not have yet a clear classification. It has been estimated that 700,000 veterans required psychological support from 1964 to 1973, making a new form of diagnostic classification necessary (Crocq and Crocq 2000). Published in 1980, the Diagnostic and Statistical Manual of Mental disorder III (DSM-III) for the first time introduced PTSD as a disorder characterized by traumatic re-experiencing, numbing, alterations of the arousal, and avoidance of the trauma-reminders (American Psychiatric Association 1980).

The understanding of PTSD has been evolving in the last forty years, driven by a growing interest of scientific research on its psychological and neurobiological basis. Beyond the historical evolution of its diagnostic criteria (for a review, see North et al. 2016), there is nowadays a consensus in recognizing the DSM-5 diagnostic criteria for PTSD (American Psychiatric Association 2013), described in the following paragraph.

3.2. Clinical features

The DSM-5 defines PTSD as a disorder developed following the exposition to one or more traumatic events. To be diagnosed, PTSD requires the presence of symptoms belonging to each of the following five symptoms criteria defined by the DSM-5 (American Psychiatric Association 2013).

- **Criterion A: Exposure to stressor.** The person must been exposed to a situation endangering his/her life, or threatening his/her physical or psychological integrity, or to sexual violence. Importantly, after several modifications through different DSM editions in decades, the DSM-5 goes beyond the early assumptions of PTSD as a war-specific disorder, by assuming that the type of exposure could be both direct and indirect, (e.g., the person witnessed the traumatic experience, or learned that a relative was exposed, or was exposed to aversive details concerning the trauma).

- **Criterion B: Intrusion.** The person persistently re-experiences the traumatic experiences in one or several ways, including unwanted intrusive memories,

nightmares, and flashbacks. This intrusive re-experiencing causes emotional distress and increased physical reactivity.

- **Criterion C: Avoidance.** The person intentionally or non-intentionally avoids trauma-related stimuli, including internal stimuli, such as thoughts, memories and feelings, and external stimuli, such as places, persons and situations.

- **Criterion D: Negative alterations of cognition and mood.** The person experiences negative feelings and thoughts, presenting at least two of the following conditions: negative affects; difficulty in experiencing positive affect; feeling isolated; dissociative amnesia; overly negative thoughts and assumptions; blame of self or other for causing the trauma; decreased interest in activities.

- **Criterion E: Alterations in arousal and reactivity.** After the trauma, the person can be on a constant hyper-vigilance state, became irritable and aggressive, develop risky behaviour, or presenting difficulties in concentrating or sleeping.

In order to fulfil the DSM-5 criteria for PTSD, the person must present symptoms for more than a month (**Criterion F**), such symptoms create distress or impairment in personal life functioning (**Criterion E**), and must not be due to substance or other pathological conditions (**Criterion H**).

Despite the DSM-5 requires the clinical manifestation of symptoms belonging to all the criterions, evidence suggests a clinical relevance for individuals showing partial (or sub-threshold) PTSD. Some persons exposed to a traumatic experience may develop only some of the required symptoms, most frequently satisfying intrusive and hyperarousal symptoms (i.e., criterion B and H, respectively). It has been showed that these patients with partial PTSD presented high levels of social and work functioning impairment, as well as suicide risk, and distress, comparable to complete PTSD (Zlotnick, Franklin, and Zimmerman 2002). Furthermore, the duration of the symptoms and the clinical manifestation could be heterogeneous and vary depending on the type of traumatic experience.

3.3. The central role of intrusive memories in PTSD

The intrusion of a distressing, traumatic past into the present is the main clinical feature of PTSD. Intrusive symptoms have received a particular interest from the scientific research, potentially representing a central hub connecting other clinical features. In the aftermath of a traumatic event, people can repeatedly experience intrusive trauma-related memories, which take various sensorial forms such as images, sounds, smells, tastes or body sensations and, less commonly, thoughts (Michael et al. 2005). While intrusive memories tend to weaken or disappear after few weeks or few months in most of people exposed to trauma, they can persist for years in people developing PTSD.

Trauma-related intrusive memories present some important specific characteristics. First, they are **involuntary and uncontrollable**: the retrieval of the traumatic experience occurs out of the person's willing. Second, they are **repetitive**: although the involuntary recalling of memories may seem an everyday common phenomenon, trauma-related intrusions are characterized by the repeated retrieval of certain sensorial aspects of the trauma. Their content is redundant, with some studies suggesting to be the moments immediately preceding the trauma or cantered around the worst moments of the traumatic experience (Holmes, Grey, and Young 2005). This redundancy is an important feature differentiating traumatic intrusive memories from everyday involuntary retrievals. Third, traumatic memories are **extremely vivid**: they sometimes take the form of flashbacks, a transient dissociative distortion of the reality. Due to the extreme vividness of the traumatic memory contents, people often experience the sensation that the event is reoccurring in the present, namely a sense of "newness", up to a total loss of connection with the self and the present in the most severe cases (Brewin et al. 2010a). Fourth, intrusive memories are accompanied by an avalanche of **strong emotional responses**. Re-living the trauma through intrusive memories elicits the emotions felt during the traumatic experience itself. Such emotions, often involving fear and terror, are accompanied by physiological reactions (Ehlers 2010).

Many external cues can trigger involuntary intrusive memories. It has been reported that newspapers and TV reporting similar events can evocate such intrusions. Interestingly, not always the cues triggering intrusive memories have a meaningful objective relationship with the trauma. Rather, intrusive trauma-related memories are thought to be triggered by cues subjectively associated with the trauma, such as the perceived physical properties of the stimuli. These associations have been reported to be out of the awareness of people, resulting

in patients with PTSD sometimes assuming no conscious direct link between the context and the intrusions (Ehlers, Hackmann, and Michael 2004).

Intrusive memories have are highly related with other PTSD symptoms, potentially representing a determinant factor. A cognitive model proposed by Ehlers and Clark (2000) theorized that intrusion symptoms are the central hub of PTSD symptomatology. Specifically, as intrusive memories provoke distress and negative emotions, people with PTSD can avoid the reminders of the trauma in order to prevent intrusions. Furthermore, intrusive memories can disrupt concentration and attention, impairing daily life functioning (Holmes et al. 2017). Confirming theoretical cognitive models, experimental research studies have found that the distress related to intrusive retrieval, the sense of re-living the trauma and the lack of context in intrusive memories are predictive of PTSD severity (Michael et al. 2005). A recent study addressing the same question through network analyses has found that intrusive symptoms are strongly connected with other symptoms' clusters, particularly avoidance, especially in the early aftermath of the trauma, but also 12 months after (Bryant et al. 2017). Due to the central role of intrusive symptoms in connecting other clusters of symptoms in the early phase of PTSD, re-experiencing symptoms may be predictive of the disorder time course, and the authors proposed that new treatments should focus on early interventions targeting intrusive symptoms in the acute phase following the traumatic experience.

4. Models of PTSD

Several cognitive and neurobiological models of PTSD have been proposed in the last decades, each of them explaining some aspects of such complex and multidimensional psychiatric condition. This chapter will propose an overview on the models and research evidence framing PTSD as (1) a memory disorder, (2) an active forgetting disorder and (3) a prediction disorder.

4.1. PTSD as a memory disorder

The idea that PTSD is a memory disorder has been proposed since long time and popular theories confer a central role to memory processes in this psychiatric condition (van der Kolk 2007). While disturbances in PTSD have been reported in a variety of memory domains (see Brewin, 2011 and van Marle, 2015) for exhaustive reviews), here the focus will be lying on a keystone memory-based model of PTSD, the **dual representation theory**, and the scientific evidence on its neural basis, mostly to be found in the hippocampus and the amygdala.

Initially formulated by Brewin, Dalgleish, and Joseph (1996) and revisited in 2010 (Brewin et al. 2010a), the dual representation theory propose that PTSD intrusive symptoms, considered as central in this psychiatric condition, arise from an incongruence between contextual and sensorial memory representations. The authors define **contextual memories** as containing traces of past experiences that can be voluntarily and consciously retrieved. Such memory traces are often abstract, declarative representations within their autobiographical context and can be manipulated and updated to respond to new situations. Contextual memories constitute the grounds for autobiographical memory and support cognitive functions such as planning, narration and communication.

By opposite, **sensorial memories,** which encode low-level sensorial and emotional representations, are involuntary retrieved by context-independent sensory features in the environmental stimuli. By assuming different names and specificities, the differentiation

between these two types of memory has been largely described in the literature (Brewin et al. 2010). Contextual and sensorial memory representations sensibly differ in their neural substrates. While sensorial memories are encoded in low-level sensory cortices, in the amygdala and in the insula, contextual memories require the activation of the medial temporal cortex (MTL), including the hippocampus, which relays the information to high-order cortical regions, by contextual information to memory representations

According to the dual representation theory, while in normal condition a sensorial representation has a correspondent contextual representation, such correspondence is compromised in PTSD. In healthy individuals, the association between contextual and sensory representations of emotional and stressful autobiographical events creates long-lasting memory traces. Importantly, sensory representations are associated to contextual representations, with the involvement of the precuneus, integrating the events and the related emotions in their contexts. Contextual integration allows disambiguating events sharing some features. A weak or absent contextual memory representation of the traumatic event, along with a stronger sensory representation, has been proposed to be the origin of intrusive symptoms in PTSD. These propositions are supported by evidence that, in parallel with involuntary perceptually and emotionally vivid memories of the trauma, people with PTSD often are not able to voluntary recover the details of the traumatic experience (Brewin et al. 2010a).

According to the dual representation model, the formation of sensory representations precedes the formation of contextual representations in time and, in PTSD, the initial intense sensory information is not copied into the hippocampal-dependent contextual memory system. Brewin et al. (2010) hypothesized that this lack of integration could be due to the effect of stress. As the traumatic event contains fundamental information to the individual's future survival, the persistence of memory traces is an adaptive mechanism. However, the distress associated to sensorial representations would prevent the event to be correctly processed by higher-order brain structures in order to contextualize the sensory memories. Another factor that could prevent the contextualization of traumatic memories is constituted by the fact that individuals with PTSD avoid trauma reminders and memories. Accordingly, the avoidance of the external stimuli that might remind the trauma would deprive subjects by the possibility to experience safe context as safe.

Stress negatively affects the **hippocampal functioning**, which plays a major role in the formation of contextual representations, and potentiates the activity of **amygdala**, which, on the contrary, is important to store sensory and emotional memory representations. Information about the context of a given event, for instance, information about locations, time, environmental and cognitive circumstances, enable the flexible representation and storage of the past, in a multimodal representation necessary for future retrieval and adaptive behaviour (Maren, Phan, and Liberzon 2013). In other words, correctly processing and contextually integrating an experience enables using these traces to predict the possibility of future similar experiences, allowing adapting behavioural responses. Animal studies have been largely focusing on the hippocampus, bringing strong evidence that hippocampal lesions compromise contextual memories encoding and retrieval. Similar findings have been found in human studies using fMRI (see Maren, Phan, and Liberzon 2013 for a review).

Fear conditioning is a straightforward paradigm to investigate contextual learning. Inherited from the pioneering works of Ivan Pavlov, this paradigm implies the association between a contextual stimulus (CS) with an aversive outcome (or unconditioned stimulus, US). Similarly to Pavlov's dog salivated when a bell rang, rodents and human beings rapidly learn associations between contextual stimuli and aversive events such as electroshocks, and they can persist in the same responses when CS is presented without any aversive outcome. Fear conditioning studies have shown that the hippocampus, but not the amygdala, is involved in encoding contextual memories (Pohlack et al. 2011; Rudy, Barrientos, and O'Reilly 2002; Zelikowsky, Bissiere, and Fanselow 2012). These studies suggested that the hippocampus is crucial for encoding and retrieving contextual memory traces, and contextual memories formed out of the hippocampus decay in time without the contribution of such subcortical structure. The hippocampus is therefore essential for linking different aspects of a memory trace, encoded by different cortical regions, in order to place them on their context and create a unified and coherent memory.

As predicted by Brewin and colleagues' dual representation theory, and compatibly with their proposed importance for contextual memories, the structural and functional integrity of the hippocampus is disrupted in PTSD. Smaller hippocampal volume in PTSD has been revealed and replicated by numerous MRI studies (see Pitman et al. 2012 for a review). Several meta-analysis including many studies with matching PTSD and control sample have shown that the reduction in hippocampal volume in PTSD is bilateral (Smith 2005) and that this reduction covariates with PTSD symptoms severity (Karl et al. 2006). Despite the

remarkable interest of these findings, the hippocampus is an extremely complex structure, with different functionally different regions. The hippocampus can be divided into different sections presenting different cytoarchitectonic and functional characteristics (Duvernoy 2005). Anatomically, the hippocampus can be divided into: (1) the Dentate Gyrus (DG), (2) the Subiculum, (3) the Cornu Ammonis (CA), which presents four different subfields (CA1, CA2, CA3, CA4), and (4) the tail. In a recent study conducted in our lab, Postel et al. (2021) investigated CA1 and a cluster composed by CA2-3/DG hippocampal subfields' volumes in individuals exposed to the Paris November 13th terrorist attack developing PTSD and not

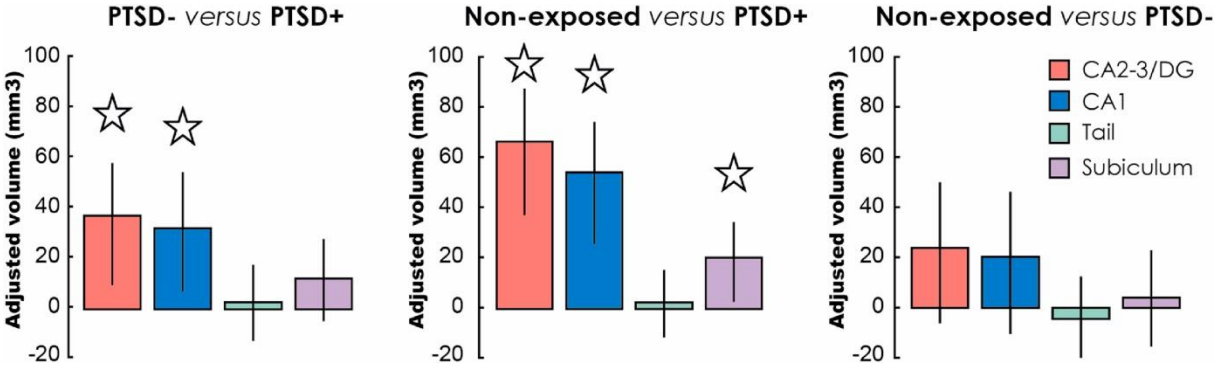


Figure 4. Group differences in hippocampal subfields between individuals developing PTSD (PTSD+), resilient individuals (PTSD-), and control group (Non-exposed). Adapted from Postel et al. (2021).

developing PTSD (i.e., resilient), and individuals nonexposed to the trauma. The authors found significant reductions in the volumes of both CA1 and CA2-3/DG in individuals developing PTSD following the traumatic experience when compared to both resilient and control individuals (see **Figure 4**).

Interestingly, this reduction in the volume of CA1 was correlated with the severity of intrusive symptoms. As CA1 is involved in enriching neural memory representations with contextual information (Barrientos and Tiznado 2016), this correlation may support the dual representation theory. The authors also found that the volume of CA2-3/DG, a region central to fear overgeneralization (Besnard and Sahay 2016), correlated with both avoidance and depression. However, the role of the hippocampus in the formation and expression of traumatic memory is far from being completely understood in PTSD, especially in the early phases following the trauma.

Individuals with PTSD also present aberrant hyperactivity of the amygdala when processing distressing and threatening stimuli (Badura-Brack et al. 2018). Similarly, amygdala's hyperactivation has been reported when individuals with PTSD were exposed to

trauma-related contents and narratives (Shin, Rauch, and Pitman 2006). The amygdala is involved in threatening stimuli, and it has been suggested that its hyperactivity in PTSD could reflect the hyper-vigilance to potential threat observed in PTSD. Compatibly with the role of hyperactive amygdala, a rodent study has found that noradrenergic augmentation in this brain structure during fear retrievals can enhance the persistence of the memory trace, hypothesizing a similar mechanism in traumatic memories in PTSD (Dębiec, Bush, and LeDoux 2011).

Although more research is needed to well characterize these mechanisms in PTSD, altogether, these findings suggest hippocampus-mediated poor contextual integration and amygdala-mediated emotional over-consolidation in trauma-related memories in PTSD. By borrowing an example proposed by (Maren et al. 2013) to illustrate the crucial contribution of the context in memories, when a fearful external cue such as a snake signals a potential danger, the hippocampus and the ventromedial PFC allow integrating the stressful stimulus in its context (i.e., where is the snake? Is it at the zoo or in the woods? Is it moving towards me?). Without contextual information, a snake is just a snake: a fearful animal endangering the individuals' safety and life.

The sensorial and emotional aspects of the threatening cue are processed by the amygdala, the parietal lobe and primary sensory areas. In PTSD, the lack of contextual information and an exaggerated emotional response may prompt the memories about the trauma assuming the form of intrusive, vivid, uncontrollable, perceptually and emotionally stressful intrusive memories characterizing the disorder.

4.2. PTSD as an active forgetting disorder

An alternative account of PTSD proposes that it could be a disorder or forgetting, rather than a disorder of memory. If, as described above, prefrontal control over memory regions contributes to suppressing intrusive memories and weakening their accessibility for future recall, then it is reasonable to investigate whether such control mechanisms may be disrupted in PTSD.

Deficits in **inhibitory control** have reportedly been observed in individuals with PTSD. People with PTSD show poorer performances in inhibitory tasks such as the Stroop

and the Go/No-go tasks (see Banich et al. (2009) for a review). Furthermore, sustained and selective attention, as well as working memory abilities, is often impaired in people with PTSD. These deficits in executive functions, defined as the control of complex goal-directed behaviour, are often driven by reductions in the activity of the PFC (Aupperle et al. 2012), a key region for inhibitory control in a variety of complex tasks, including control over intrusive memories (see the paragraph **Neural bases**). Despite evidence indicating the existence of an executive and inhibitory deficit in PTSD, only a few studies have up to now investigated SIF in this disorder.

In one of the first behavioural studies addressing this research question, Catarino et al. (2015) proposed that a deficit of inhibitory control may underlie intrusive memories in PTSD. Studying a cohort of 18 individuals with PTSD and 18 trauma-exposed individuals without PTSD, the authors implemented a naturalistic variant of the TNT task, in which participants learned associations between objects and aversive scenes and were asked to suppress such scenes when objects were presented during the TNT phase. To test the SIF effect, after the TNT, participants performed a recall test, consisting in a verbal description of the scenes associated with each of the object presented during the previous phase. Results revealed that the PTSD group showed impaired SIF. While resilient individuals showed decreased identification and details in the description of the suppressed scenes when compared to non-suppressed scenes, this effect was not observed in the PTSD group. Interestingly, the number of retrieved details about suppressed scenes correlated negatively with PTSD symptoms' severity, suggesting that higher degrees of SIF corresponded to lower disease severity.

In a recent TNT fMRI study, Sullivan et al. (2019) investigated the neural correlates of memory suppression in trauma exposure and PTSD. The researchers found that, regardless of their clinical diagnosis, trauma-exposed participants had significantly reduced right MFG activations when compared to the healthy control group, suggesting that the simple exposition to a traumatic experience can affect the recruitment of the memory control system, but not PTSD itself. However, these results should be interpreted in the context of several study limitations. It should be noted that the sample size was constituted by 35 trauma-exposed participants, of which 16 had PTSD and 19 did not, and 13 nonexposed participants, weakening the strength and power of statistical comparisons. Also, the nature of the trauma varied across participants. Furthermore, as highlighted above (see the paragraph **Neural bases**), analyses of connectivity are better suited to characterize inhibitory control mechanisms than univariate fMRI analyses. On the opposite, a magnetoencephalography

(MEG) revealed that, during suppression attempts, PTSD patients were unable to downregulate signatures of sensory long-term memory traces in the gamma frequency band, compared to control participants with the same trauma history but without PTSD (Waldhauser et al. 2018).

In a recent research study, we questioned whether the brain mechanisms normally supporting the suppression of intrusive memories are disrupted in PTSD, and whether the preservation of such mechanisms could, on the contrary, be associated with resilience in the aftermath of an acute and severe traumatic exposure (Mary et al. 2020). In this study, the largest study investigating memory suppression PTSD, a group of 102 participants exposed to the Paris November 13th terrorist attacks and a matching control group composed by 73 nonexposed individuals performed the TNT task, in order to suppress neutral and inoffensive intrusive objects associated with word. Crucially, within the trauma-exposed group, 55 participants suffered from complete or partial PTSD (Zlotnick et al. 2002) and 47 participants showed no impairment after trauma. Functional connectivity analyses revealed that resilient and nonexposed participants exhibited decreased coupling between brain control regions (including the MFG, the superior frontal gyrus, the inferior frontal gyrus and the cingulate cortex), and memory retrieval regions (including the hippocampus, the parahippocampal cortex, the precuneus and the fusiform gyrus) during intrusive trials when compared to both nonintrusive trials and resting state. This pattern, consistent with an increase of inhibitory control when people suppress intrusive memories, was not observed in individuals with PTSD (see **Figure 5**, on the top). DCM analyses (see the paragraph **Dynamic Causal Modelling**) confirmed that absence of increased top-down downregulation over memory regions, orchestrated by the dlPFC, when people with PTSD counteract intrusive memories entering consciousness (see **Figure 5**, on the bottom). Please, see the full-text study of Mary et al. (2020) in the **ANNEX**). These fascinating results provide strong evidence in favour of a novel interpretation of PTSD as an active forgetting disorder, pointing out at the disruption of memory suppression as a risk factor for developing PTSD following a traumatic experience. In parallel, the preserved functioning of the brain mechanisms underlying memory suppression may constitute a factor promoting positive adaptation and resilience following a traumatic experience.

Despite these recent encouraging results and the fact that influential models of active forgetting theorized a disruption of suppression mechanism in PTSD (Anderson et al. 2016; Anderson and Hulbert 2021), a comprehensive neurobiological and cognitive model of such

disruption is still missing. Further research is needed to integrate findings proving alterations in the brain mechanisms underlying memory suppression and classic theories focusing on memory disorders in PTSD. It is conceivable that these two accounts are in fact not mutually exclusive.

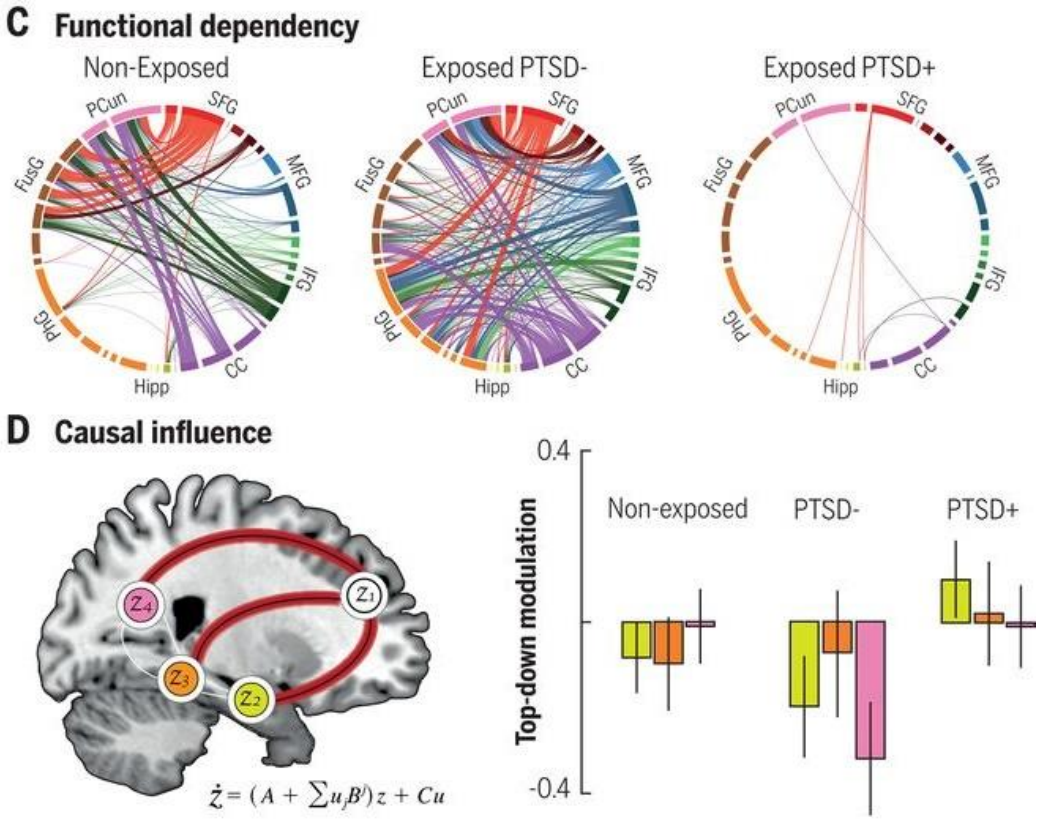


Figure 5. On the top, decreases in coupling between control and memory regions in intrusive relative to nonintrusive condition, as revealed by PPI analyses. On the bottom: differences between DCM coupling parameters during intrusive and nonintrusive trials. The lack of decreased coupling during intrusive condition in PTSD was confirmed by DCM analyses, which showed no significant differences in top-down downregulation during intrusive and nonintrusive conditions in the PTSD group. Adapted from Mary et al. (2020).

4.3. PTSD as a prediction disorder

Some researchers have recently proposed mathematically-informed models, suggesting that PTSD is a disorder of predictions (Gagne, Dayan, and Bishop 2018; Homan et al. 2019; Seriès 2019). Accordingly, surviving to extremely negative and threatening life

events can lead people to shift their behaviour towards the avoidance of contexts that could lead to similar experiences. This would results in alterations of the homeostatic equilibrium, characterized by an imbalance between approach and avoidance behaviours (Stein and Paulus 2009). This relatively simple hypothesis have been related with some of the symptoms of PTSD: hyperarousal would result from an upregulation of the avoidance, and anhedonia and substance use would arise from a downregulation of approach (Stein and Paulus 2009).

Avoidant behaviour can also prevent the correct processing of the traumatic event and lead to aberrant predictions about the probability that the event will happen again. The traumatic event can disrupt the normal associative learning, strengthening the associations between neutral stimuli such as places, people or sounds, and threatening outcomes. While these associations normally weaken or disappear in time, they persist in PTSD, possibly as a consequence the avoidant behaviour. Indeed, avoiding neutral stimuli previously associated with the trauma would prevent the possibility to face again these reminders, possibly experiencing them as safe and in this way updating and perhaps extinguishing their associations with threatening outcomes (Seriès 2019).

In this context, re-experiencing plays a crucial role for the maintenance of the disorder. In a recent theoretical model, Gagne, Dayan, and Bishop (2018) hypothesized that a traumatic event generates a large discrepancy between the expected and the actual outcome (i.e., a prediction error, PE), which, given its salience for personal survival, triggers the interior re-experiencing of the event and its antecedents. Repeatedly replaying the trauma via intrusive memories, flashbacks and rumination would strengthen the associations between the neutral characteristics of the trauma antecedents (i.e., places, people, sounds, etc.) with threatening outcomes, resulting in aberrant predictions about the probability that the event would happen again when exposed to such reminders. Furthermore, re-experiencing can favour the over-generalization of fear. When re-experiencing the trauma, negative value would be conferred to situations and actions sharing features with the trauma reminders, at more or less concrete level, generalizing aberrant predictions (Gagne et al. 2018). Thus, in simple terms, this account of PTSD proposes that, along with persistent re-experiencing, an associative learning deficit would compromise the ability to correctly predict aversive events, in turn exacerbating avoidant behaviour.

A growing body of empirical evidence supports this novel interpretation of PTSD. By building mathematically-informed models of associative learning, two recent studies have

found increased sensitivity to surprising information, that is, an increased weighting of the PE, affecting learning processes in PTSD in both neutral (Brown et al. 2018) and threatening (Homan et al. 2019) experimental contexts. This over-sensibility to PE correlated with increased symptoms' severity and increased amygdalar activity when processing unexpected negative events.

These latter two studies applied computational methods to model how beliefs about the outcomes are formed and updated. A relatively novel approach to computational disorders, namely computational psychiatry, allows quantifying individual differences in terms of learning, expectations, and PE. Still at its infancy, computational modelling of is a promising tool for testing hypotheses on the pathological cognitive and brain mechanisms of PTSD (Seriès 2019). The next chapter will present a detailed overview of the theoretical framework of computational psychiatry, as well as an overview of the most important methods useful to investigate the hidden cognitive and brain dynamics underlying healthy and pathological functioning.

5. Computational Psychiatry

5.1. Towards a computational definition of mental disorders

The idea that the brain is a complex machine solving computational problems dates back to Alan Turing's pioneering works (Turing 1950). It is, however, in the latest decade that this idea has found its applications in psychiatry. In this framework, the brain is a statistical organ continuously building and updating complex models of the world to generate hypotheses about its functioning and to test them against the sensory evidence (Friston et al. 2014). If the brain is a problem-solving machine, then the origin of atypical functioning observed in psychiatric conditions would arise from aberrant computations.

The need for a new approach to psychiatry is principally due to the lack of diagnostic tools able to establish observable and reliable markers of the mechanisms underlying observable symptoms. Indeed, modern psychiatry is mostly based on the categorization of observable symptoms into clusters (as in the DSM and in the International Classification of Diseases, ICD) to characterize complex and often phenomenologically overlapping disorders. These classifications have been found to be poorly reliable and predictive of the patients' clinical trajectories in time (Stephan and Mathys 2014). Different physiological mechanisms could be affected at different degrees across patients, resulting in the lack of clinical interpretability of diagnostic labels. A crucial assumption is that diagnostic categories are consequences, and not causes of the psychopathology. In other words, psychiatric symptoms are the observable consequence of hidden, unobservable pathophysiological and psychopathological mechanisms (Friston, Redish, and Gordon 2017). The aim of this novel computational approach in psychiatry is to fill the gap between these hidden pathological mechanisms and the related symptoms, in order to predict the individual clinical evolutions and promote personalized treatments.

Computational psychiatry characterizes the computations that the brain performs and characterizes how such computations could be affected in psychiatric disorders (Adams, Huys, and Roiser 2016). Several theory-driven and data-driven computational approaches have been implemented (for a review, see Stephan et al. 2015). **Generative models** constitute a promising computational instrument to study psychiatric disorders, and they are gaining popularity. Generative models allows building and testing hypotheses-driven mathematically-informed models on how observable variables (e.g., symptoms) could be generated by hidden (e.g., physiological or cognitive) mechanisms. Generative models describe hypotheses on how observed data are generated, in a Bayesian framework. Formulated by Thomas Bayes in 1763, the **Bayes' theorem** simply describes the probability of an event based on prior knowledge. In its simplest applications, the Bayes theorem aims to compute the probability that an event *A* occurs given an event (or previous evidence) *B*.

Generative models specify hypotheses on how hidden, unobservable states $\boldsymbol{\theta}$ modelled through a model $\boldsymbol{m}$ probabilistically map into real measurements $\boldsymbol{y}$ obtained through neuroimaging or behavioural paradigms. These models adapt the Bayes theorem to combine the **likelihood function** $\boldsymbol{P(y|\theta, m)}$, which expresses the formal probabilistic mapping from model parameters $\boldsymbol{\theta}$ to observations $\boldsymbol{y}$ and the **prior knowledge** on the parameters' probability under a specific model $\boldsymbol{P(\theta|m)}$. A process called **model inversion** (or model inference) allows computing the **posterior probability distribution** of the parameters $\boldsymbol{\theta}$, given the data under the model $\boldsymbol{P(\theta|y, m)}$, as follows:

$$\boldsymbol{P(\theta|y,m) = \frac{P(y|\theta,m) * P(\theta|m)}{P(y|m)}} \qquad \textbf{Eq. 1}$$

Where $\boldsymbol{P(y|m)}$ refers to the model evidence, or marginal likelihood, encoding the likelihood to obtain the observed data under the specified model. Importantly, all these probability distributions are Gaussian, described by their sufficient statistics, whose variance indicates their uncertainty.

The inversion of a generative model consists in the estimation of both model evidence $\boldsymbol{P(y|m)}$ and posterior parameters $\boldsymbol{P(\theta|y, m)}$. The computation of the model evidence requires dealing with complex integrals, often impossible to solve analytically. For

this reason, during the model inversion an approximation of the model evidence is required. Even if some different methods exist, such as Markov chain Monte Carlo, the **Variational Bayes approximation of the Free Energy** has been the most implemented in computational neuroscience, showing high consistency and low computational costs. Introduced on thermodynamics in the seventies by Richard Feynman (Feynman 1998), and lately adapted to neuroscience by Friston (2009, 2010), in the context of generative model inversion the free energy is an upper-bound on the (log-) model evidence. Thus, the variational inversion of the model consists in the optimization of the posterior parameters that minimize the negative free energy, that is, the parameters that maximize the model evidence.

As a schematic example of generative models, let's imagine a measure of brain activity, for example using fMRI or electroencephalography (EEG), for which we should model the dynamics governing such activity (see Figure 6). A model of ionic channel can be built and tested against real neuroimaging data, in a first step called the forward model. The forward model describes hypotheses on the hidden ionic channels' dynamics underlying and generating the detected brain activity. Model inversion (or inference) consists in the estimation of the posterior distribution of the parameters that are modelled as causing the fMRI or EEG observed signals, i.e., in this example, the ionic channels' activity (Frässle, Yao, et al. 2018; Stephan et al. 2016).

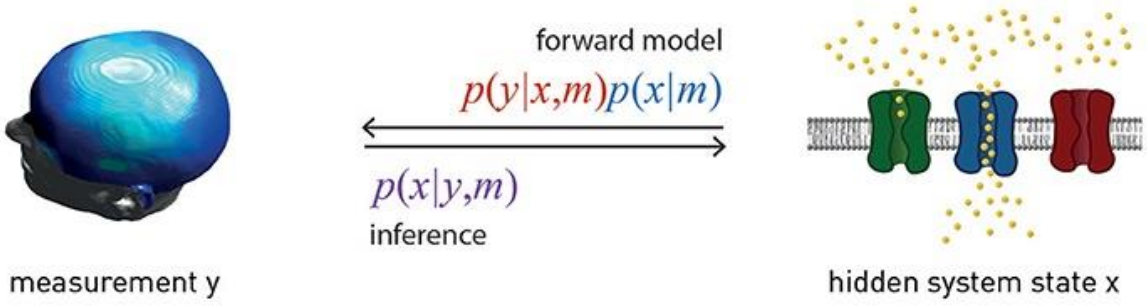


Figure 6. A schematic example of generative models. Adapted from Stephan et al. (2016).

Computational psychiatry has an unprecedented potential for understanding psychiatric disorders, as well as a for translational psychiatry research. Ideally, hypotheses-driven generative models of the mechanisms underlying mental disorders can be built and applied to brain or behavioural functioning in individual patients, in order to detect physiologically and/or cognitively subgroups belonging to the same diagnostic category and predict individual diagnostic evolutions and response to different treatments (see Erreur ! Source

du renvoi introuvable. and Frässle et al. 2018). Generative models are receiving a growing attention in the last years, and they have been mainly implemented to characterize the computations underlying human behaviour and brain functioning in psychiatric disorders. In the next two sections, attention will be paid to the computational methods developed to model behaviour and brain dynamics, respectively, as well as on the issue of model selection performed to identify the best generative model from a set of candidate models.

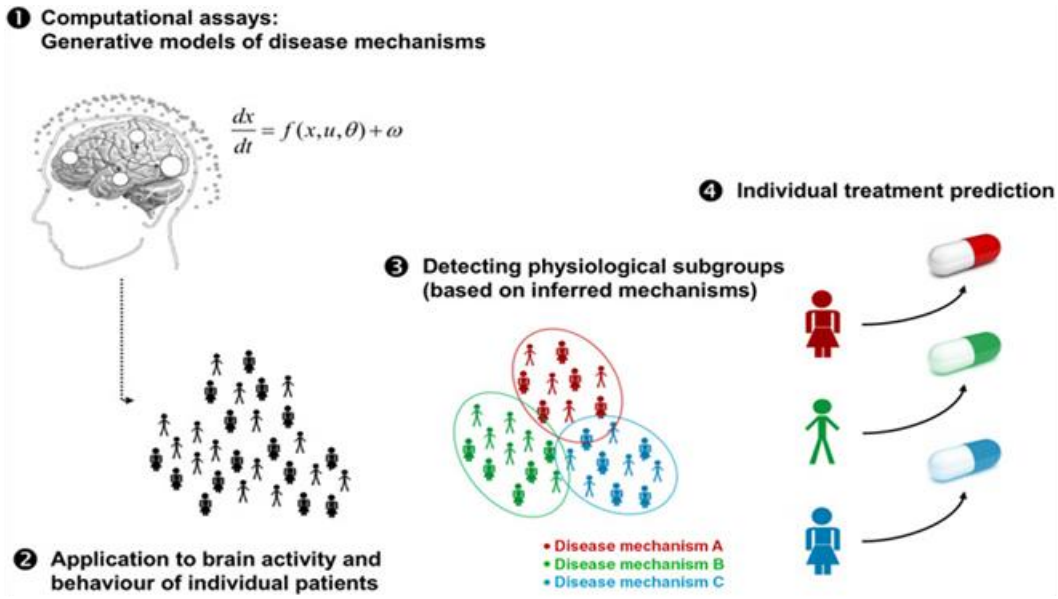


Figure 7. Schematic example of generative models. Adapted from Stephan et al. (2016)

5.2. Bayesian modelling of human behaviour

According to the theoretical framework of computational psychiatry, the brain is not a passive filter of the reality but, on the contrary, an active organ generating hypotheses about the states of the world to adapt behaviour to external demand and actively interacting with its environment (Adams et al. 2016).

The **Bayesian brain theory** proposes that the brain combines prior expectations and sensory data to produce beliefs or predictions to explain the rules governing the environment. According to his Bayesian framework, we can imagine an agent and the world as two separate statistical units interacting and exchanging each other. There are hidden states ruling the world, which are not accessible to the agent, who needs to capitalize on the sensory evidence to produce an optimal model to infer the hidden states. This model has to be as accurate as

possible, in order to infer the discrepancy between the predicted and the real world and minimize surprise. When, following a novel encounter with sensory evidence, the predicted world does not coincide with the real world (i.e., there is PE), beliefs about the hidden states are updated (see Figure 8).

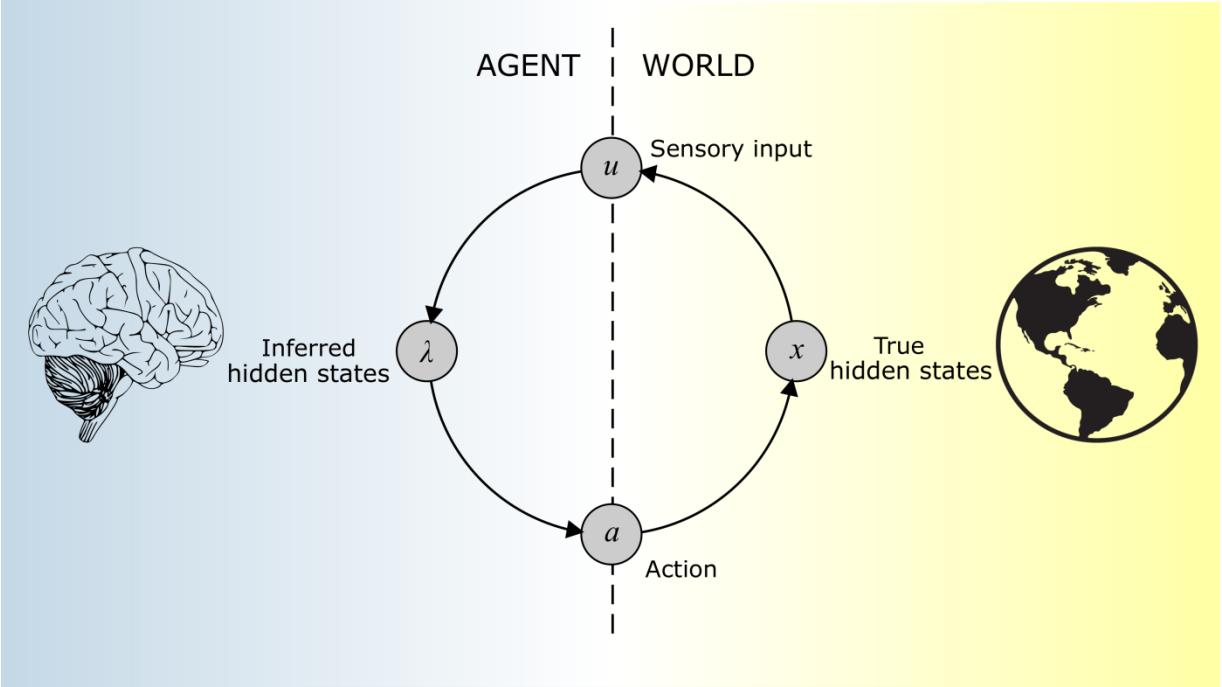


Figure 8. Schematic representation of the Bayesian brain theory.

Crucially, sensory information, as well as beliefs, is somehow uncertain. Like a generative model, the Bayesian agent integrates uncertain priors and likelihood probability distributions, defined by their mean and precision (i.e., the inverse of the variance) to produce posterior beliefs about the world. The optimal integration of prior expectations about the environmental hidden state and the likelihood of observing a sensory input given these prior expectations to produce an accurate model of the world is assumed to follow the rules of the Bayes theorem (see above). This process aims to the form and update a model of the world by optimizing its evidence world (Friston et al. 2014).

The interaction between the agent and the world is bidirectional: the environment shapes the agent's internal beliefs and predictions and, on the same time, the agent modifies its own environment with active goal-directed behaviour and decision-making (see Figure 8). The **active inference** framework suggests that the action can change the environment, thus contributing to the maximization of the internal model evidence (i.e., the minimization of the free energy). For instance, behavioural responses can favour the equilibrium between the

external and the internal world by actively adapting the real world to the expected world or by biasing the sampling of sensory evidence towards the ones fulfilling the expectations (Friston et al. 2010).

Under the computationally psychiatry framework, the brain is considered a generative model of the world (Friston et al. 2014). A recent technical approach, named **meta-Bayesian modelling**, propose to model such generative models through generative models (Daunizeau, den Ouden, et al. 2010; Daunizeau, den Ouden, et al. 2010). This approach aims in "*observing the observer*", that is, building computational models on how the individuals build their own computational models of the world. Accordingly, at the subject (i.e., the observer) level, a Bayesian observer forms a "**perceptual model**" to transform sensory evidence into an internal model of the world, which, from a Bayesian perspective, encodes the posterior beliefs about the environment. The perceptual model is built to understand the hidden states ruling the environment and predict the future encountering of sensory signals. The "**observation model**" (or "response model") aims describing the mapping from inferred hidden states to behavioural outcomes. At the experimenter (i.e., the observer's observer) level, only sensory inputs and behavioural responses are known; in other words, only the consequences of the internal perceptual models are observable.

The experimenter aims building generative models to infer the (unknown) generative models of the experimental subjects. To do this, the experimenter inverts the subjects' generative models. The inversion of the response model allows mapping from behavioural responses to their causes (i.e., the beliefs). However, this inversion implies an inversion of the perceptual model, to map from sensory inputs to the subjects' beliefs. This approach requires building hypotheses-driven computational models coding hypotheses on how the subjects form both perceptual and observation models (Daunizeau, den Ouden, et al. 2010; Daunizeau, den Ouden, et al. 2010).

This approach has become a popular solution for inferring subjects' beliefs on a great variety of experimental settings, often associated with records of brain activity. Evidence has shown that the human beings form **hierarchical beliefs** when facing to uncertain environments, in a range of domains covering probabilistic mapping between stimuli and outcomes (de Berker et al. 2016; Iglesias et al. 2013), spatial attention (Vossel et al. 2014), social learning (Diaconescu et al. 2017; Siegel et al. 2019), and more. Depending on the

computational model implemented and on the experimental procedure, the definition of beliefs' hierarchy is different.

All the aforementioned studies implemented the Hierarchical Gaussian Filter (HGF, see the **Experimental studies** part for a detailed description) to infer internal beliefs. According to this general model developed by Mathys et al. (2011) under the meta-Bayesian approach, at least three levels of beliefs with ascending degrees of complexity and abstraction are distinguishable. On the first, lowest level, the individuals predict the immediate future sensory states, basing on their prior and likelihood distributions. This level is the sigmoidal transformation of the second level, which encodes beliefs about the presence of contingency in the environment which may evolve in time (i.e., its volatility). The third level encodes beliefs about the volatility of those changes in time (i.e., meta-volatility). Importantly, each level evolves through the precision-weighting of ascending hierarchical PE, which are fundamental for beliefs' updating across levels. Thus, the confrontation with sensory evidence discrepant with the internal model of the world would drives the updating of beliefs.

Beyond the mathematical details of this specific computational model, described in details below, it is important to know that this hierarchy on beliefs, especially in an uncertain environment, have been reportedly found. Some neuroimaging studies have confirmed the hierarchical organization of beliefs and PE in the brain (Diaconescu et al. 2017; Iglesias et al. 2013). Furthermore, signal processing in the brain reflects these hierarchical dynamics. In an outstanding study, Rao and Ballard (1999) have found a hierarchical **predictive coding** in the visual cortex. The authors proposed a model of visual processing in which high-order visual areas attempt to predict the activity of lower-order areas via forward, top-down connections. Errors when these predictions are tested against sensory evidence (i.e., PE) would result in ascending signalling triggering the updating of high-order predictions.

Thus, the brain appears to actively making predictions about its own functioning and its relationship with the external world, online learning through trial-and-errors, and adapting its predictions to maximize their accuracy. This novel Bayesian modelling approach has the merit of having placed back cognitive (computational) processes on the centre of the psychiatry research interest. Beliefs can be altered in psychiatric disorders. For instance, Bayesian modelling has contributed to demonstrate that adults with autism spectrum disorder form aberrant beliefs on the volatility of the environment (Lawson, Mathys, and Rees 2017), as well as that individuals with major depression present impaired use of past rewards to make

decisions (Rupprechter et al. 2018), that delusions in psychotic patients may represent rigid beliefs without sufficient evidence (Corlett and Fletcher 2014), and more.

Aberrant beliefs may be the core of most of psychiatric disorders, as these deficits in computations can induce emotional and somatic symptoms and promote the maintenance of adaptive behaviours (Moutoussis et al. 2018). Generative models of the human behaviour and cognition are a valuable tool for the understanding of psychiatric disorders, by linking these hidden computations to the related symptoms, and promoting strategies in psychotherapy that may remediate these pathological internal models of the world.

5.3. Dynamic Causal Modelling

Another class of generative models aims to the comprehension of hidden mechanisms regulating the brain activity signals measured through neuroimaging modalities such as fMRI and EEG. Developed by Friston, Harrison, and Penny (2003), DCM is a method allows inferring hidden neural states underlying Blood-oxygen-level-dependent (BOLD) or electrophysiological activity time-series, such as the connectivity between neural populations and their synaptic strength. DCM presents two important characteristics:

1. **DCM is dynamic**: By using differential equations, its aim is to describe unobservable time-dependent dynamics. As the other classes of generative models described above, DCM requires building hypotheses-driven models of the possible causes of the observed data.

2. **DCM is causal:** It allows inferring causalities in the connectivity between different neural populations, that is, how the neural dynamics underlying the activity a brain area directly cause the activity of another brain area via their synaptic connections. Furthermore, the causal effect of external inputs and experimental manipulations on synaptic changes can be modelled.

This method provides estimates of the strength of the synaptic connections between different brain regions (Stephan et al. 2010). Contrarily to functional connectivity, which is a synthetic representation of unobservable brain dynamics, effective connectivity corresponds to the parameters of a generative model trying to explain those unobservable phenomena (Friston 2011). This difference is crucial: going beyond the correlational nature of functional

connectivity analyses, DCM unambiguously models the neurobiological processes underlying the excitatory or inhibitory connections among neuronal populations.

DCM entails building a biologically plausible **forward model** describing hypotheses on the causes of the neural dynamics provoking regional BOLD changes, including hypotheses on the causal connectivity among regions and the causal effect of experimental manipulations, and the subsequent model inversion (see Figure 9).

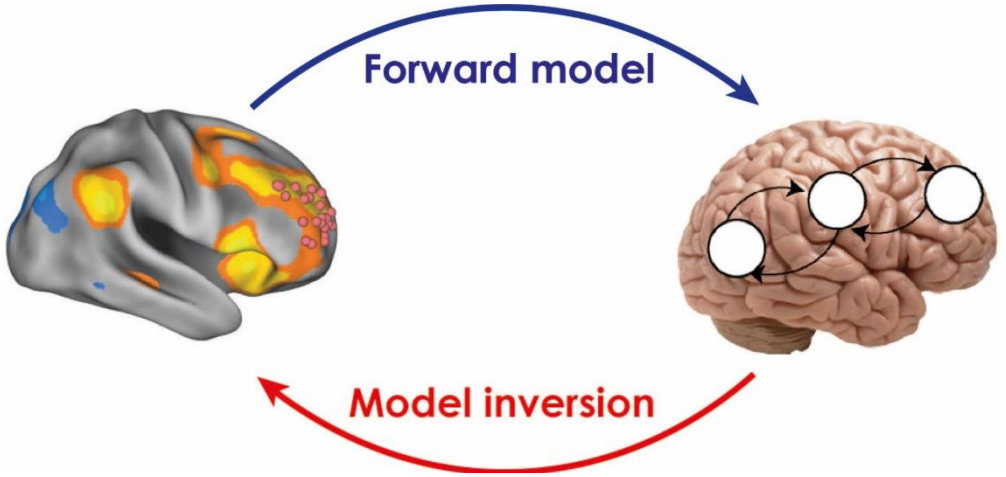


Figure 9. Schematic representation of the generative processes underlying DCM.

The forward model is the combination of a neural model and a haemodynamic model (for fMRI applications). The **neural model** describes hypotheses on the hidden neural dynamics $\frac{dx}{dt}$ of a neural system, represented by their activity *x* at time *t*, basing on the following general bilinear state equation:

$$\frac{dx}{dt} = \left(A + \sum_{j=1}^{m} u_j B^{(j)}\right) x + Cu \qquad \textbf{Eq. 1}$$

Given *m* known inputs, this equation describes the hidden dynamics $\frac{dx}{dt}$ with three matrices:

- The A matrix describes hypotheses on the endogenous connectivity between brain regions in the absence of experimental modulations.
- The B matrix describes the j^{th} modulatory input u_i operating on the intrinsic connectivity between brain regions. This matrix is the most important, because it encodes hypotheses on how experimental manipulations causally modify the

connectivity from a region A to a region B (i.e., their coupling). The modulation is an additive change to the A matrix intrinsic connectivity.

- The matrix C describes hypotheses on how extrinsic inputs (e.g., experimental manipulations) drive the *activity* of one brain region.

An additional matrix D can be added to this equation to model the nonlinear modulations of the connectivity between two brain regions by the activity of a third region (Stephan et al. 2008). As a Bayesian generative model, the estimation of DCM coupling parameters require specifying prior distributions for each of these matrices. These priors reflect the empirical knowledge about the range of plausible values that these parameters can have.

While DCM has been widely used also for EEG data, here the focus will be deliberately laid on its application to fMRI data. Put simply, fMRI allows measuring BOLD signal local changes, which are usually interpreted as changes in brain activity (for a detailed review, see (Heeger and Ress 2002)**.** In DCM for fMRI time series, the neural model is convoluted to the **haemodynamic model**, a biophysical model that transforms the neural activity to measurable signal. This complex model, derived from an extension of the "Balloon model" (Friston et al. 2000), translates the neural activity into changes in terms of vasodilatation signals, blood cerebral flow and deoxyhaemoglobin concentration.

The haemodynamic model allows converting the hypotheses encoded by the neural model into a synthetic BOLD signal, to estimate the neural model's parameters that maximize the similarity between the real and the predicted BOLD. As for other classes of generative models, the Variational Bayes inversion of the DCM model furnishs posterior parameter distributions ($\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{m}$) for each of the components of the bilinear state equation (i.e., matrices A, B and C). These posterior parameters are Gaussian distribution whose mean corresponds to the maximum a posteriori probability of the parameter and its variance is the parameters' covariance. Posterior parameters are interpretable as the coupling parameters of the intrinsic causal connectivity (matrix A), the extrinsic modulation of this connectivity (matrix B), and the effect of extrinsic inputs (matrix C). In summary, these parameters represent the effective connectivity.

As for computational modelling of human behaviour, the model inversion schema aims to the minimization of the negative free energy, by estimating the parameters that

minimize the divergence between the predicted brain activity and the observed brain activity. The model inversion also output the evidence of the model $\boldsymbol{P(y|m)}$, indicating the probability of observing the data under a specific DCM model (Stephan et al. 2010).

Although some applications have been proposed to model whole-brain or large scale neural populations (Frässle, Lomakina, et al. 2018), DCM normally requires the selection of a few a priori regions of interest (ROI) to investigate their effective connectivity. Furthermore, the definition of prior distributions is very important for the model inversion. DCM represents an unparalleled tool for studying brain connectivity, and its application in psychiatric disorders have been shown to be highly beneficial in the understanding of behavioural and cognitive dysfunctions (Heinzle and Stephan 2018). However, DCM requires strong a priori hypotheses and rigorous experimental procedure: several possible pitfalls and precious rules have been reviewed by Stephan et al. (2010). Another important point to acknowledge is that the model evidence of one model is not informative of the goodness of fit of such model: building and comparing different models encoding different hypotheses is fundamental for characterizing complex phenomena in the brain. In the next section, a brief overview on the methods for comparing different generative models will be presented.

5.4. Bayesian model selection and averaging

In the framework of computational psychiatry, a crucial advantage is the possibility to compare generative models encoding different competitive hypotheses on the hidden mechanisms underlying the cognitive or brain functions of interest. Bayesian model selection (BMS) is a technique allowing comparing (log-) model evidence in order to select the most probable one given the data. In a very simple view, BMS consists in inverting generative models attempting to explain the observed (behavioural or physiological) symptoms to obtain their likelihood $\boldsymbol{P(y|m)}$, and then compare this likelihood to determine the model best explaining the observed data (see **Figure 10**). In the context of the DCM, for example, one may want to test alternative hypotheses on the directional nature of the connectivity between two regions during an activation task (e.g., top-down or bottom-up connectivity) to select the hypothetical model that most likely generated the data.

Although some alternative exist, such as Bayesian Information Criterion and Akaike Information Criterion , evidence shows that the free energy approximation of the model

evidence is the most reliable index of the model likelihood to compare different models (Stephan et al. 2017). Contrarily to these alternatives, this method aims to find the model with the optimal balance between accuracy (i.e., the probability of observing the data *y* given the model *m*) and complexity (i.e., the difference between priors and posteriors and the covariance among the parameters of a model).

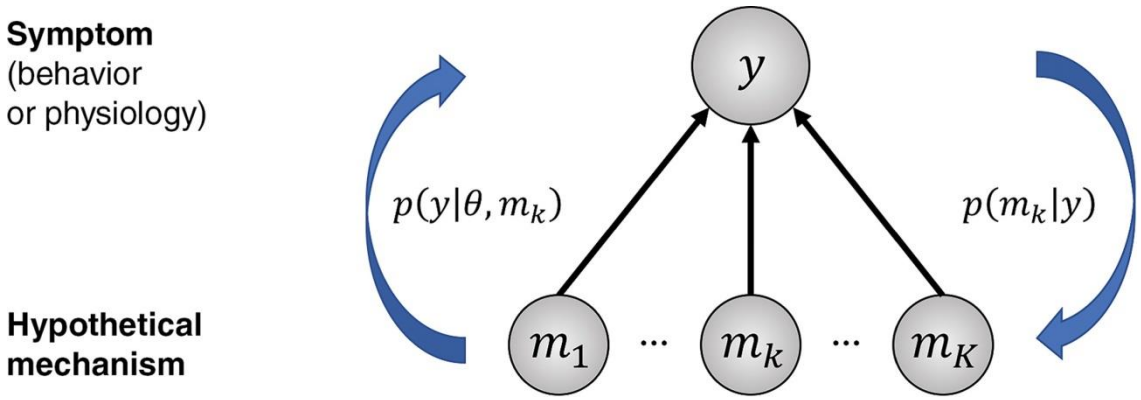


Figure 10. Schematic illustration of the BMS framework. Adapted from Stephan et al. (2017).

Developed by Stephan and colleagues (Stephan et al. 2009), BMS has been applied to DCM group studies to either infer differences in the structure of the effective connectivity or to select the best model to make inferences at the parameters' level (e.g. on coupling parameters). By using the free energy approximation of the model evidence of a set of competitive models, BMS computes the probability of a model being more likely than the other model in the model space (i.e., the exceedence probability, XP). A recent development of this by Rigoux and colleagues (Rigoux et al. 2014) computes the Bayesian Omnibus Risk (BOR), which quantifies the probability that group differences in model frequencies might be due to chance. XP and BOR can then be used to compute the protected exceedence probabilities (PXP), which quantify the probability of a model to be more frequent then the others in the model space, above and beyond chance. Importantly, random-effects BMS (RFX-BMS) allows making inferences on the characteristics of the population to which the subjects belong. Furthermore, in the context of DCM, this novel method also tests the hypothesis that different groups are drown from different populations presenting structural differences in the effective connectivity.

An interesting application of RFX-BMS is to compare families of models. Computational models can be aggregated into families of models presenting some common

characteristics. For example, in DCM, a model family can hypothesize top-down connectivity from a brain region X to a set of potential target areas, while an alternative family can hypothesize that these top-down connections originate from a distinct region Y. In this context, RFX-BMS would allow establishing, across models, whether the connectivity between these regions is most likely originating from region X or Y.

When there is no evidence for a particular model being the most likely within a family of models sharing some features, Bayesian Model Averaging (BMA) for DCM studies allows to summarize family-specific coupling parameters. These coupling parameters are averaged based on their evidence. Briefly, for each participant *s* belonging to the group, the averaged parameters across the models m within the family f_D, $P(\theta_{s\in g} \mid Y, m \in f_D)$, are computed by weighting the participant's posteriors for each model m in the family (i.e., $P(\theta_s \mid y_s, m)$ by the posterior probabilities that participant *s* uses model *m* (i.e., $P(m_s \mid Y_g)$:

$$P\left(\theta_{s\in g} \mid Y_g, m \in f_D\right) = \sum_{m\in f_D} P\left(\theta_{s\in g} \mid y_{s\in g}, m\right) P\left(m_{s\in g} \mid Y_g\right) \qquad \textbf{Eq. 2}$$

where Y_g is the dataset of the whole group g, containing data for each participant in the group, $y_{s\in g}$ (Penny et al. 2010) .

In summary, a computational approach to psychiatry allows testing and selecting hypotheses on the causes generating symptoms and cognitive and brain dysfunctions in psychiatric disorders. This approach constitutes the methodological framework of the current study. As computational psychiatry is a recent discipline, despite theoretical propositions of PTSD as prediction disorder, a relatively few studies have nowadays attempted model the pathological mechanisms underlying PTSD under a computational approach. No studies have until now attempted to link prediction and memory control disorders in PTSD under a computational approach.

6. Context of the current research study

6.1. REMEMBER

On the evening of November 13th 2015, a series of coordinated jihadists terrorist attacks hit Paris and Saint-Denis, in France. After three kamikaze bombings near to the *"Stade de France"* in Saint-Denis, during the football match France-Germany, the streets and terraces of the 10th and 11th *arrondissement* of the French capital have been the target of multiple terrorists' shootings. A hostage-taking started at the *"Bataclan"* theatre, turning out into a massacre. These terrorist attacks caused the dramatic loss of 131 lives and more than 300 persons were injured.

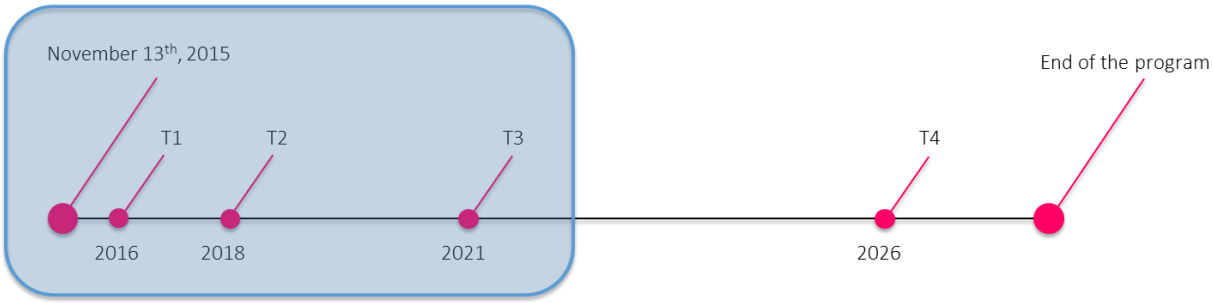


Figure 11. Longitudinal time points of the 13-Novembre research program. Highlighted in blue: the longitudinal phases of the REMEMBER project.

The November 13th terrorist attacks represent a trauma at both individual and societal levels. Following the attacks, a large research program named "***13-Novembre***", aiming to understand the construction and the evolution of the individual and collective traumatic memories, has been set up (https://www.memoire13novembre.fr/). This longitudinal research project promotes the collaboration of neuroscientists, historians, sociologists, anthropologists, and public health and law researchers, and the recollection of multidisciplinary data on different time points over 12 years (see **Figure 11**).

The main investigation of this program, named *"Étude mille"* ("1000 study"), aims recording filmed testimony about the terrorist attacks through structured interviews in a sample of 1000 voluntary participants with different degrees of exposition to the traumatic event. These participants can be: directly exposed (C1); locals and neighbours of the attacks' places (C2); living in the suburban areas of Paris (C3); living in provincial cities of France (C4). These participants will be recalled for four follow-ups in 10 years (see Figure 12). This branch of the program, based in Paris, investigate the formation of evolution of individual and collective memories and narrations from both qualitative and quantitative points of view.

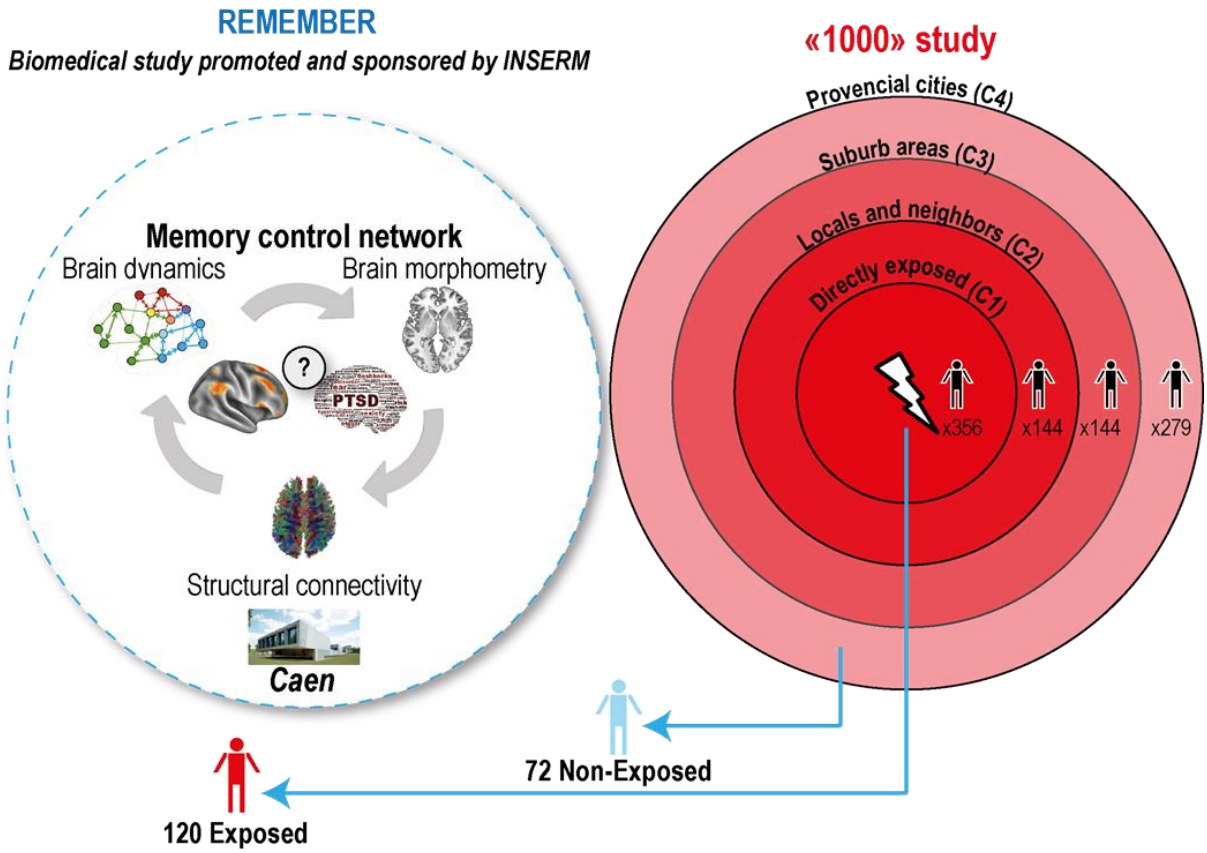


Figure 12. Participants of the 13-Novembre research project at the phase 1. One thousand voluntary participants were included to the "Étude mille" and 192 participants were included in the REMEMBER project. Figure realized by Pierre Gagnepain, adapted under permission.

In the context of this research program, a biomedical research study, named ***REMEMBER***, is conducted in Caen, France, and aims to understand the structural and functional brain markers associated to the development of PTSD following the traumatic event. *REMEMBER*, standing for "REsilience and Modification of brain control network following novEMBER 13") is a longitudinal project involving three phases in 6 years (see Figure 11). At time 1 (T1) the study included 72 participants not exposed and 120 participants

directly exposed to the Paris terrorist attacks and at time 2 (T2) the study included 70 nonexposed and 107 exposed participants. Crucially, the core aim of the REMEMBER project is to understand why some people develop PTSD while other do not, after a traumatic event. The main hypothesis is that PTSD may be characterized by functional alterations in the brain network normally supporting memory suppression and structural alterations in the hippocampus. In order to study the brain mechanisms of memory suppression, participants performed the TNT task (see the paragraph **The Think/No-think task**) while recording fMRI brain activity. Resting state fMRI activity was also recorded. Other MRI modalities were acquired, including structural MRI to investigate grey matter volume and thickness, diffusion tensor imaging (DTI) to investigate the integrity of white matter fibre tracks, and a high resolution sequence to investigate hippocampal subfields. Thus, by including different brain imaging modalities, this research project allows isolating different markers of maladaptive response to trauma and resilience, as well as to understand how these markers evolve and cover psychopathology in time. The first results of *REMEMBER* have shown that people developing PTSD following the 13th November terrorist attack exhibit altered suppression of intrusive memories (Mary et al. 2020, see also the paragraph **PTSD as an active forgetting disorder** and the **ANNEX**) and alterations in the hippocampal subfields volumes (Postel et al. 2021, see also the paragraph **PTSD as a memory disorder**). Developed in the context of *REMEMBER*, the current PhD work attempts to understand in a first study the contribution of aberrant predictive processing to the expression of active forgetting disorder in individuals with PTSD. We took advantage of the aforementioned methods of computational psychiatry to address this important research question and proposes the concept of **predictive control**, rooted in the relationship between the brain's predictive and control mechanisms. In a second study, we focus on the longitudinal changes in predictive control, and further quantify parallel alterations of the hippocampus, to understand how these markers, representing distinct accounts of PTSD, may drive the clinical trajectories of exposed individuals.

6.2. PTSD as a predictive control disorder

People developing PTSD following a traumatic event overestimate the probability of aversive events. These aberrant predictions could drive the avoidance of trauma reminders to prevent threat and reduce stress. Due to their vivid contents and emotional load, intrusive trauma-related memories could be the target of avoidant behaviour. However, little is known

about the mechanisms underlying the prediction and the avoidance of intrusive memories, and the different mechanisms that support their suppression.

Inhibitory control of behaviour, cognition and motor action can be reached via early predictive (or proactive) mechanisms and late reactive corrections. Computationally, proactive control would be driven by the beliefs (or predictions) about the optimal amount of control to deploy, and reactive control would add further fine-grained control to suppress remaining prediction error signals (Braver 2012; Jiang, Heller, and Egner 2014).

It has been proposed that memory control can also be reached via predictive and reactive mechanisms (Anderson et al. 2016; Levy and Anderson 2012a). However, previous TNT studies were not able to disentangle these two dynamics. TNT studies usually reported brain activations or connectivity after contrasting the intrusive condition with the non-intrusive condition. This procedure does not allow inferences about proactive and reactive dynamics. While memory control nonintrusive trials could be interpreted as purely driven by predictive dynamics, intrusive trials would contain both predictive and reactive signals, making the interpretation of the observed brain activity ambiguous. Bayesian modelling of behaviour allows inferring subjects' beliefs about the probability that the upcoming cue will trigger an intrusion and, accordingly, predictive control would be driven by such beliefs. At the same time, PE, represented by the difference between the real outcome (i.e., intrusion or non-intrusion) and prior beliefs would drive the reactive adjustment of memory control.

First study

In the first study of this thesis ***"Predicting the future to suppress the past",*** we considered the link between predictive and active forgetting disorders in PTSD. We hypothesized that individuals with PTSD form aberrant beliefs about the probability of experiencing intrusive memories. As we hypothesized that this prediction disorder is a general dysfunction rooted in the brain predictive system, we used the TNT task with neutral, non-threatening stimuli. This allows us to put exposed and nonexposed participants in the same footage, not biasing our study. To test this hypothesis, we applied Bayesian modelling to infer participants' beliefs during the NT trials, and the hidden parameters governing these beliefs.

We also hypothesized that the paradox of the harmful avoidance of traumatic memories observed in PTSD may arise from the disrupted balance between predictive and

reactive memory control dynamics towards the first. This would lead to cognitive over-anticipation of intrusive memories and the parallel failure in blocking intrusions when they occur in PTSD. We propose that these dynamics arise from the disruption of the memory control brain system at a broad level, not uniquely confined to traumatic memories. To test this hypothesis, we used trajectories of beliefs and PE as modulators of the effective, causal connectivity between the control system (composed by the anterior and the posterior MFG) and the memory system (composed by the rostral hippocampus, caudal hippocampus and the precuneus), as indexes of predictive and reactive control, respectively. We then investigated the relationship between the balance of predictive and reactive control and both trauma-related and trans-diagnostic clinical features of PTSD, and further applied circular statistics to assess the imbalance between predictive and reactive mechanisms. This study is currently under review on a peer-reviewed scientific journal.

Second study

In the second study of this thesis, ***"Plasticity of memory control and hippocampal circuits forecast remission from PTSD"***, we used a longitudinal design to investigate the predictive value of memory control and hippocampal disorders in forecasting future PTSD symptoms' evolution. PTSD has long been considered as a memory disorder characterized by alterations in the hippocampus.

We have shown that PTSD is linked to a disorder of inhibitory control of intrusive memory, characterized by the imbalance between predictive and reactivity. We investigated whether remission from PTSD was associated with changes in predictive and reactive memory control and hippocampal subfields' (HS) volumes. We collected data on predictive and reactive control, as in the first study, and on hippocampal subfields' volumes 6 to 18 months at time 1 (T1), and 30 to 42 months at time 2 (T2), after the traumatic event. At T2, 18 participants remitted from PTSD. We hypothesized that this group was characterized by increased reactive control of intrusive memories and plasticity of the hippocampus.

We also hypothesized that evolution in memory control, especially, improvements in reactive control, and hippocampal plasticity may forecast future decreased symptoms' severity. A third data recollection round involved clinical assessments using PCL-5. We tested the relationship between both the evolution of predictive and reactive control and the

evolution of CA1 and CA2-CA3-DG volume and the future evolution of symptoms' severity, between T2 and T3. This study is currently under preparation for the submission to a peer-reviewed scientific journal.

EXPERIMENTAL RESEARCH STUDIES

7. First study

Predicting the future to suppress the past after trauma

Giovanni Leone[1], Charlotte Postel[1], Alison Mary[1,2], Florence Fraisse[1], Thomas Vallée[1], Fausto Viader[1], Vincent de La Sayette[1], Denis Peschanski[3], Jaques Dayan[1,4], Francis Eustache[1], Pierre Gagnepain[1]*

[1] Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, GIP Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France.

[2] Neuropsychology and Functional Imaging Research Group (UR2NF), Centre for Research in Cognition and Neurosciences (CRCN), UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium.

[3] Université Paris I Panthéon Sorbonne, HESAM Université, EHESS, CNRS, UMR8209, Paris, France.

[4] Pôle Hospitalo-Universitaire de Psychiatrie de l'Enfant et de l'Adolescent, Centre Hospitalier Guillaume Régnier, Université Rennes 1, 35700 Rennes, France.

***Corresponding author email**: pierre.gagnepain@inserm.fr

Pierre Gagnepain

GIP Cyceron, Boulevard Becquerel

14074, Caen, France

Tel.: +33 (0)2 314 701 59

Abstract

Aberrant predictions of future threat lead to maladaptive avoidance in individuals with post-traumatic stress disorder (PTSD). How this prediction disorder influences the control of memory states orchestrated by the dorsolateral prefrontal cortex is unknown. We combined computational modeling and brain connectivity analyses to reveal how individuals exposed and nonexposed to the 2015 Paris terrorist attacks formed and controlled beliefs about future intrusive re-experiencing during a memory suppression task. Exposed individuals with PTSD formed aberrant beliefs and used them excessively to control hippocampal activity. When this predictive control failed, the prediction-error associated with unwanted intrusions was poorly downregulated by reactive mechanisms. This imbalance was linked to avoidance, but not to general disturbances such as anxiety or negative affect. Conversely, trauma-exposed resilient and nonexposed individuals were able to optimally balance predictive and reactive suppression. These findings highlight a new pathological mechanism of PTSD rooted in the relationship between the brain's predictive and control mechanisms.

Keywords: inhibitory control, predictive control, memory suppression, Bayesian modelling, effective connectivity, computational psychiatry, hippocampus

Introduction

Individuals with post-traumatic stress disorder (PTSD) avoid traumatic reminders in order to anticipate threat[1,2] and reduce distress. Their perception of the future may have changed in the aftermath of the traumatic experience[3,4]. Bayesian models of the brain[5] provide a natural solution to understand the impairment of these predictive processing[6]. More specifically aberrant associations may arise between safe environmental cues and threatening outcomes[5,7], thereby compromising their ability to accurately predict aversive events[8]. This prediction disorder exacerbates the avoidance of trauma reminders[1], preventing the extinction or updating of the traumatic engram. The impact of this prediction disorder on the control of the re-experiencing of unintentional flashbacks or intrusive memories (i.e., cardinal symptom of PTSD[9]), however, is unknown.

In a recent study, we suggested that the persistence of intrusive memories in individuals with PTSD may be rooted in a generalized dysfunction of the inhibitory control system that normally regulates unwanted memories[10]. In this study, 102 participants who had been exposed to the November 2015 Paris terrorist attacks, as well as 73 nonexposed individuals, learned a series of neutral words paired with images of objects, and were later instructed to suppress the unwanted re-experiencing of intrusive memory images involuntarily triggered by the word reminder cue. During this suppression phase, we recorded brain activity using functional magnetic resonance imaging (fMRI), and participants were asked to report the presence or absence of intrusions at each trial. Crucially, exposed participants were divided into two subgroups: individuals with PTSD symptoms, and resilient individuals who had successfully dealt with the trauma. Resilient individuals exhibited a decrease in functional coupling between control and memory brain networks during the experiencing of intrusive memories, compared with both nonintrusive and resting-state conditions. This pattern is consistent with an increase in inhibitory (i.e. negative) coupling during suppression of intrusive memories. Dynamic causal modeling (DCM) analyses confirmed that this decrease in coupling reflected top-down mechanisms orchestrated by the right dorsolateral prefrontal cortex (DLPFC)[11]. In memory regions involved in the persistence of the trauma, such as the hippocampus and precuneus (PC)[12], this controlled down-regulation of intrusive memories was severely compromised in individuals with PTSD, whose brain dynamics did not differ between the intrusive and nonintrusive conditions.

These findings highlight a fundamental role of memory control mechanisms in the development of PTSD in response to trauma, but tell us nothing about the origin of their disruption and the potential contribution of hidden computations underlying predictions of intrusive memories. Cognition, motor responses and memories can be controlled by an early proactive mechanism that biases attention according to goals, and additionally corrected during a late reactive process[13,14]. Interestingly, these processes are captured well by Bayesian models that incorporate the dynamic adjustment of predictions based on previous experiences and the use of prediction error (PE) to modulate the future need for control and its correction[15]. The prediction-based dynamic adjustment of the forthcoming need for control reflects a form of *predictive control* that critically depends on the DLPFC[16]. We hypothesized that the inhibitory control of memories also relies on predictive inferences, and that the interaction between predictive and control processes is central to understanding the pathogenesis of PTSD.

We can assume that estimated probabilities of intrusive re-experiencing based on prior encounters (i.e., beliefs) are aberrantly prioritized in individuals with PTSD, such that control resources are allocated to a form of predictive avoidance that overrides online memory signals. For instance, individuals with PTSD may not only avoid situations for which they anticipate flashbacks, such as certain places or times of the day, but may also use this expectation to proactively alter conscious thoughts. Alternatively, the reduced inhibitory control in individuals with PTSD may be limited to reactive processes targeting the online emergence of intrusive memories, given their hypersensitivity to PE[5] which may reduce the control resources available and inhibitory coupling. In the context of a memory suppression task, prior exposure to successive reminders influences the belief that an undesired memory will emerge into consciousness while processing the upcoming cue. Critically, exaggerated predictive control, reduced reactive control, or a combination of the two may explain our previous observation that the brain connectivity markers of memory suppression are disrupted in individuals with PTSD (see Fig. 1b)[10].

In the current study, we tracked these hidden computations during the think/no-think (TNT) memory suppression task using meta-Bayesian modeling[17] and analyzed their impact on the underlying connectivity markers of memory control. We applied this analysis to the same subgroups with (PTSD+; n = 55) or without (PTSD-; n = 47) PTSD following exposure to the terrorist attacks in Paris on 13 November 2015, and the same nonexposed participants (n = 73)[10] (see Methods). We submitted trial-by-trial computations of beliefs about upcoming

intrusions and resulting PE to a DCM analysis to explore their influence on effective connectivity between the inhibitory control system and memory target regions. We focused this analysis on the right anterior and posterior middle frontal gyrus (MFG)[10,18], two core nodes of the inhibitory control system , and tested their relative contribution to belief-driven and PE-driven control. We tested the influence of these two distinct control hubs on two memory regions that are central to the establishment of traumatic memory: the hippocampus, distinguishing between its rostral and caudal parts[10] and the PC.

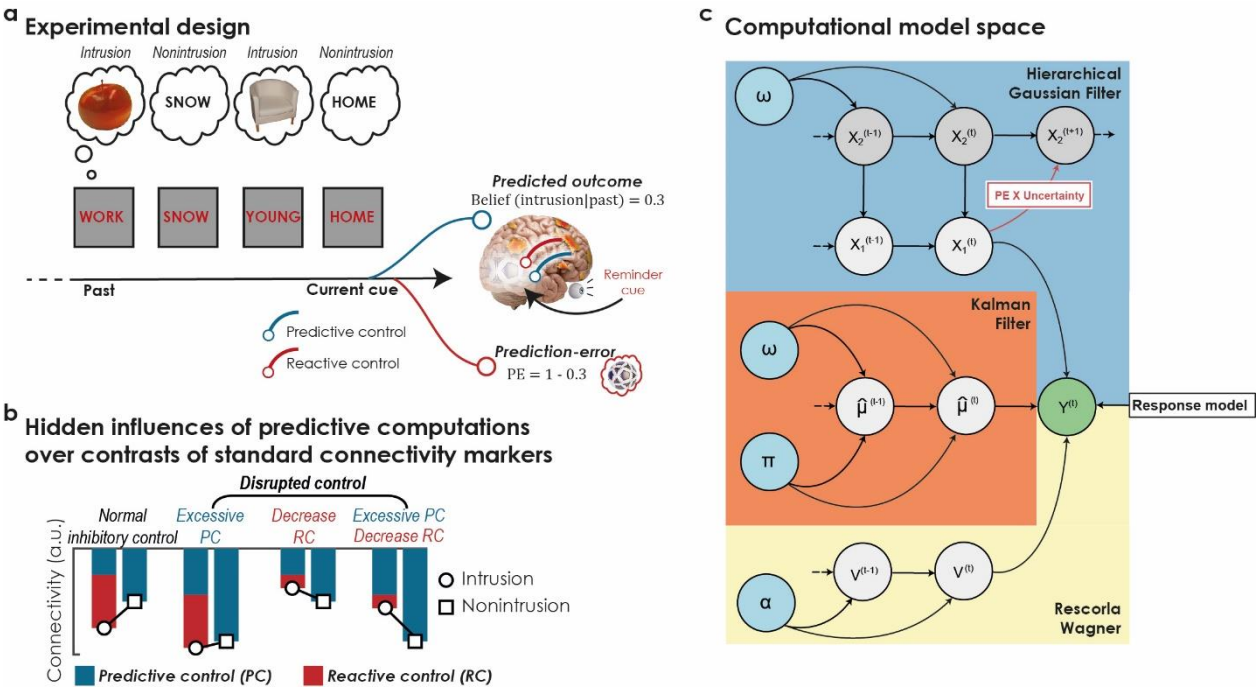


Figure 1. Design and computational models. (**A**) After learning word-object pairs, participants performed a memory suppression task in which they were asked to prevent the memory of the images associated with the cue words from entering awareness. They then rated the presence or absence of intrusive memories during suppression attempts. The estimation that an upcoming cue will trigger an intrusive memory (i.e., belief) can be inferred from previous encounters, providing an adaptive advantage in the form of the deployment of optimum memory control and proactive prevention of memory retrieval (i.e., *predictive control*). Reactive control is engaged when intrusive memories unexpectedly cross the proactive gate, resulting in a prediction error (PE) that triggers additional inhibition and updating of future expectations. It should be noted that recall cues (i.e., think items) are not displayed here (see Method). (**B**) Toy example. Standard contrast analyses of intrusive and nonintrusive cues cannot identify the contribution of these critical computational quantities on the disruption of the connectivity markers of inhibitory control. (**C**) Computational model space. Binary intrusion ratings across the suppression task were fed into computational models to track belief formation across the suppression task. In the two-level hierarchical Gaussian filter (HGF; pale blue panel), beliefs are hierarchical and dynamically weighted by uncertainty. The perceptual parameter ω regulates the speed of belief adjustment throughout the task. The Kalman filter (KF; pale orange) also includes dynamic belief updating, which is regulated by two free perceptual parameters, π and ω, encoding belief reliability and uncertainty, but it does not assume hierarchical beliefs. The Rescorla-Wagner model (RW; pale yellow) is a simpler non-hierarchical model

with a fixed, participant-specific learning rate α. The response model describes the log-probability of the behavioral outcomes (i.e., intrusion or nonintrusion rating) given beliefs through a beta density function. These trial-wise log-probabilities are used to compute model accuracy.

Results

Computational modeling

To track beliefs about upcoming intrusive memories, we applied three distinct models of increasing complexity (Fig. 1C): 1) the Rescorla-Wagner (RW)[19], which postulates that trial-by-trial PE updates belief at a fixed learning rate; 2) the Kalman filter (KF)[20], in which the updating of belief relies on a dynamic (i.e., not fixed) learning rate, shaped by additional trial-by-trial uncertainty weighting of PE, assuming that such uncertainty is constant and the learned environment not volatile; and 3) the two-level hierarchical Gaussian filter (HGF)[21] which, like the KF, assumes that the learning of belief is a dynamic process based on uncertainty, but further assumes that the environment is volatile, and which involves the hierarchical embedding of beliefs. Note, however, that the two-level HGF model can also be interpreted as a Kalman filter operating at the (logit-transformed) contingency level as opposed to simply the outcome level like our current implementation of the KF model.

We built three distinct source models to map intrusion beliefs onto outcome probabilities. Each of these models assumed different sources of beliefs, in order to establish their accuracy in predicting the outcome. The *state* source model assumed that participants formed beliefs based solely on the trial history. The *item* source model assumed that beliefs were based solely on the history of each specific word-object pair (see Fig. 2A), disregarding overall trial-by-trial history. The *combined* source model assumed that the combination of state and item (precision-weighted) beliefs improves prediction accuracy (see Eq. 14 and Fig. 2A). These three source models mapped beliefs onto binary ratings through a beta function, with a free parameter estimating the accuracy of this mapping (see Methods). Model accuracy was computed using the negative log-likelihood of the choice probability for each of the nine computational models (HFG-state, HFG-item, HFG-combined, RW-state, RW-item, RW-combined, KF-state, KF-item, KF-combined).

Model validation

We performed different simulations to determine whether our model produced valid and reliable outputs. Intrusion ratings decreased across blocks of trials in the TNT task[10]. We first performed model falsification[22] to evaluate whether our computational models could generate this expected pattern of behavioral responses across a wide range of simulated model parameters. This analysis is reported in detail in the Methods, but briefly, consisted in simulated synthetic beliefs from 200 virtual participants using the above-mentioned models, and repeated the virtual experiment 100 times using perceptual parameter randomly drawn from a Gaussian priors distribution tailored to match our own data (to sample plausible parameters), resulting in 20000 simulations for each of the nine computational models. Then, synthetic beliefs were map into binary rating which were averaged across repeated sampling and summarized as intrusion proportion across the 4 artificial TNT sessions (see Fig. 2a). Second, we tested for each model whether we could recover the simulated trajectories of beliefs, and whether these trajectories were distinguishable among competing source models. We fit synthetic binary data generated with the same, as well as competing, source models (i.e. state, item, and combined), and compared the resulting trajectories to simulated ones using correlation. Results revealed we could confidently recover the true generated trajectories among competitor source models for HGF, but not for RW or KF models (see Fig. 2b). Third, we use the same logic to verify the reliability of the model selection criterion for identifying the true generative model within a set of competitive source models, and ensure that this selection is not biased in favor of one particular model[22,23]. This procedure, known as model recovery, consists in simulating data with one specific model and then comparing the predictive performances (i.e. model accuracy) of a set of different models. This analysis confirmed that the comparison between these three source models was not biased for HGF (Fig. 2c). However, the probability of recovering the true model was confounded with competing source models for RW and KF (Fig. 2c). Fourth, we performed parameter recovery analyses[23], to ensure the reliability and meaningfulness of estimated model perceptual parameters. Results of these analyses, reported in detail in Methods (see also Fig. 2d), indicated that parameter recovery was modest for the HGF model and poor for RW or KF.

In summary, in the current experimental setting, only the HGF model produced valid trajectories of intrusive beliefs, which accurately simulate the behavioral pattern, and reliably and truthfully distinguishing beliefs formed on the basis of the trial or item history, or a combination of both memory sources. We therefore used HGF to track trial-by-trial variations

in beliefs about the potential re-experiencing of upcoming intrusive memories attached to a cue word and the resulting PE, and investigate the influence of these estimates on brain control mechanisms using connectivity analysis. The perceptual parameter (ω) for this model is a participant-specific constant indicating the speed at which these beliefs are changing. We then tested whether our model was sufficiently powered to detect changes in this parameter. To test this, we simulated and recovered parameters for two distinct synthetic groups using an effect size in a range of our data (i.e. the average difference in perceptual parameter between groups) and then performed statistical tests to detect group differences in this simulated data set. The statistical power to detect group difference on the model perceptual parameter (corresponding to the frequency of significant test in this simulated data sets) was 90% for HGF-item, 10% for HGF-state. This suggest that this perceptual parameter can be confidently recovered from intrusive beliefs and compared between groups when it is derived from the item structure, but not from task state (note, however, that the outcomes of the following analysis of connectivity are independent, and not related to this perceptual parameter; see Supplementary Fig. 1). Regarding the ω parameter computed for item beliefs, the PTSD+ group expressed significantly slower beliefs updating than the nonexposed group, $t(122) = -2.10$, $p = .037$, bootstrapped 95% CI [-.59, .06], and a trend compared with the PTSD- group, $t(99) = 1.82$, $p = .072$, bootstrapped 95% CI = [-.58, -.04]; although this effect was significant when the bootstrapping of the mean is considered). No differences were found between the nonexposed and PTSD- groups, $t(113) = 0.15$, $p = .880$, bootstrapped 95% CI [-.22, .27], item belief updating. Compared to healthy individuals, individual with PTSD were less prone to shift their beliefs about a particular item after they failed to control it and suppress the associated intrusion.

Source of intrusion beliefs

To determine the memory source of intrusive beliefs (i.e. state, item, or combined), we performed Bayesian model selection (BMS) and compared the accuracy of the three source models at the population level. This analysis revealed that the combined model (protected exceedance probability, PXP = .999) outperformed the other two source models (Bayesian omnibus risk; see Methods, BOR = 0). The probability that the same model would optimally explain data in all three groups $P(H_{F=} \mid y)$ was .996 (see Methods).

Taken together, these findings suggest that in all three groups, beliefs about the experiencing of memory intrusions across suppression attempts 1) spread according to a two-level hierarchy that took volatility of belief uncertainty into account, 2) were driven by a flexible and dynamic learning process updated by PE, and 3) originated from the merging of recent *meta-memories* about their control performance that derived from both trial history and item-specific memories, as observed in other forms of cognitive control[24].

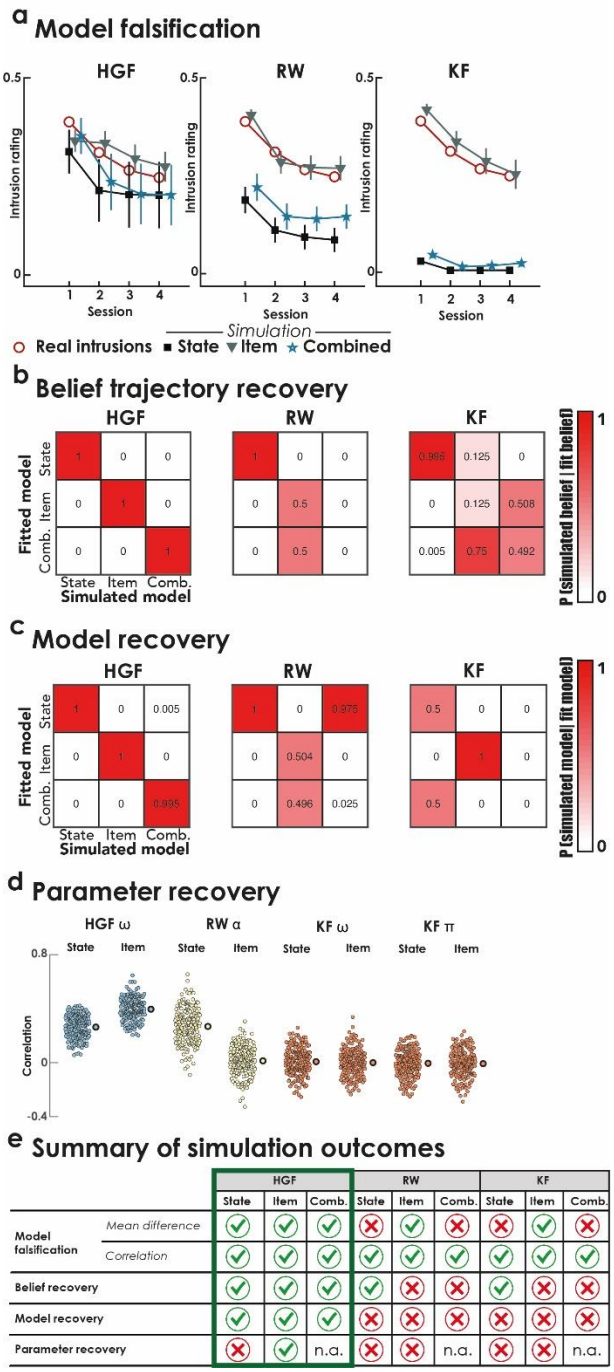


Figure 2. (**A**) Model falsification. In order to test the models' generative performance (i.e., the model ability to generate plausible data), we generated synthetic intrusion data for each model, simulating 200 virtual

participants for which we repeated the simulations 100 times (so 20000 simulations in total). We reported the session-wise mean trajectories of real intrusions rating (empty red circles) and simulated intrusions data, under HGF, RW and KF models, for both state (black squares), item (gray triangles), and combined (blue stars) sources model versions. Error bar represents 95% confidence intervals of the virtual participants' distribution. **(B)** Belief recovery. Inversion matrix reflecting the confidence that the belief fitted by a given model was the model that most likely has generated those beliefs. **(C)** Model recovery. Inversion matrix reflecting the confidence that the best fitting model has generated the data **(D)** Parameter recovery. Correlation between fitted and simulated model parameter for each virtual participant and each model. The large dot at the right of each distribution represents the mean correlation across virtual participants. **(E)** Summary of simulation outcomes.

Computational Dynamic Causal Modelling

For each cue word, our combined HGF2 computational model provided an estimate of the participant's hidden belief that the cue would trigger an intrusive memory, as well as an estimate of the discrepancy (i.e., PE) between the expected and experienced outcome (see Figs. 1c and 2a). We then investigated the influence of these estimates on brain control mechanisms, using DCM. We distinguished predictive mechanisms engaged to suppress intrusion beliefs from reactive mechanisms related to the additional demand of controlling the error induced by intrusive memories. For instance, if a cue was associated with an intrusion belief ($\hat{\mu}_1^{(t)}$) of 0.3, then the presence of an intrusion ($y^{(t)} = 1$) would require additional PE control of 0.7 ($PE^{(t)} = y^{(t)} - \hat{\mu}_1^{(t)}$, see Fig. 1a). These quantities were used as parametric modulators of the inputs (i.e., stick function) modulating the top-down coupling between control and memory systems. It should be noted that we focused this analysis on positive PE (PE+) to specifically isolate reactive control associated with suppression, and discarded negative PE associated with the absence of control demands during nonintrusive cues. However, parametric modulation of belief was performed for all cues.

We built 42 DCM models, which could be divided into three families expressing different hypotheses on the involvement of these computations. The first family, corresponding to our main hypothesis, assumed that these computations influenced top-down control. A second family tested the influence of these computations on bottom-up connections. A third family, in which the modulatory stick function of suppression trials was not parametrically modulated, tested the absence of influence of these computations on top-down control (i.e., no-computation models). Each of these families included reciprocal hypotheses about the role of the anterior MFG (aMFG) and posterior MFG (pMFG) in predictive and reactive control

(see Fig. 3a). Half the models were assigned to the predictive or reactive influence of the pMFG and aMFG, and the other half to the opposite relationship. These six subfamilies therefore each contained seven models describing the possible combinations of modulation pathways between MFG and target regions (see Fig. 3b). Target regions included the rostral hippocampus (rHIP) and caudal hippocampus (cHIP), as well as the ventral portion of the PC (see Methods for the definition of volumes of interest and timecourse extraction). In addition to these 42 models testing our main hypotheses, we included a null model family hypothesizing an absence of controlled modulation (see Fig. 3a).

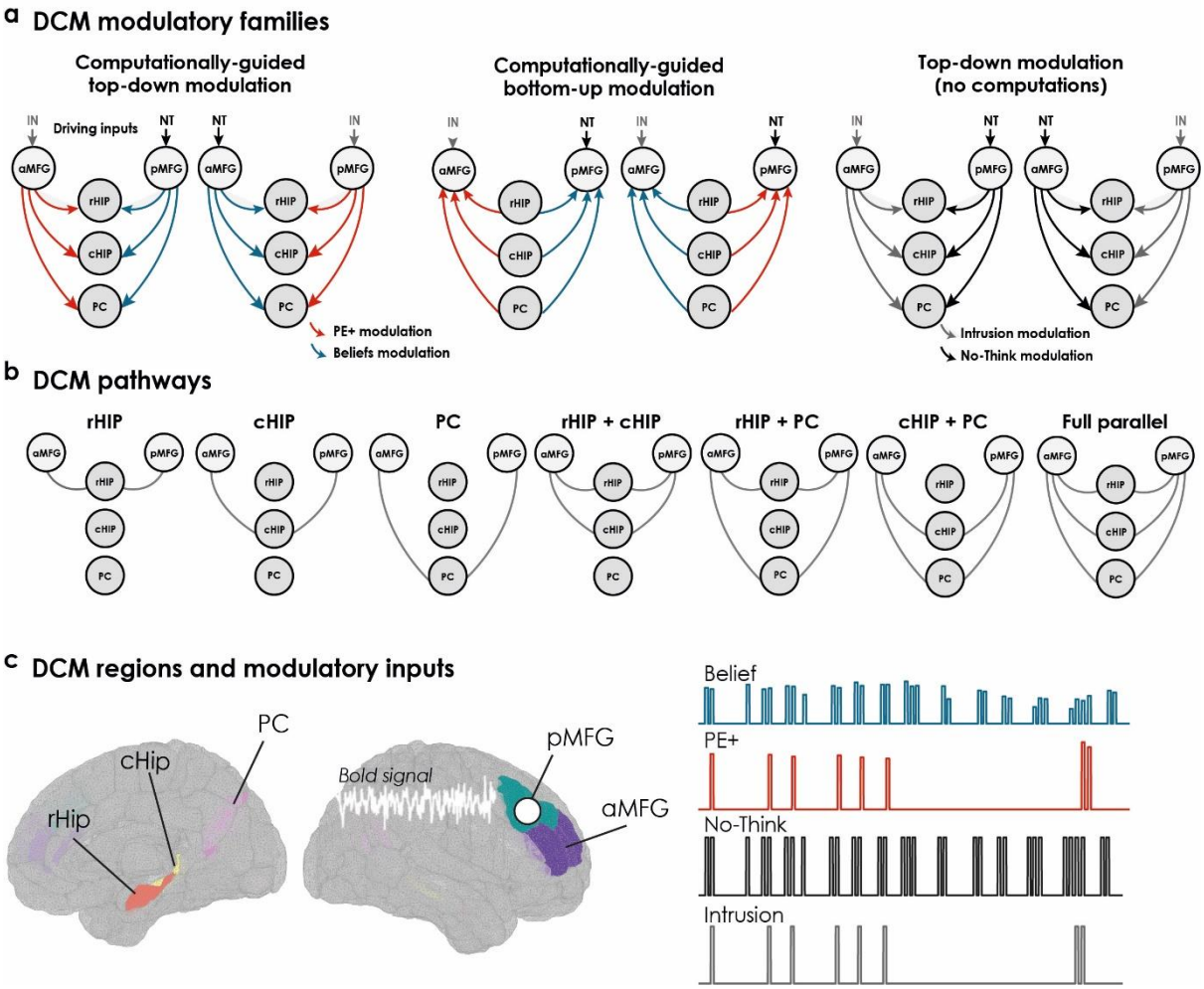


Figure 3. DCM models. (**A**) Families expressing different hypotheses on the involvement of intrusion beliefs and PE+ computations in modulation of the coupling between control regions (anterior and posterior middle frontal gyrus, MFG) and memory target regions, including the rostral hippocampus (rHIP), caudal hippocampus (cHIP), and precuneus (PC). It should be noted that null models were also estimated, but are not shown here. (**B**) Pathways capturing the seven possible connections between control and target regions. (**C**) Left panel shows the regions of interest used for DCM analysis. Right panel provides an illustration of the modularity inputs influencing the connectivity between brain regions.

Combined influence of anterior and posterior MFG during control

First, we investigated whether beliefs and PE+ effectively modulated the causal influence of MFG on memory regions across all groups. In other words, we wanted to know whether predictive and reactive control mechanisms could explain the top-down coupling between these regions during motivated forgetting. Accordingly, the 14 models assuming a top-down modulation of control by belief and PE+ (i.e., first family), were compared with the models belonging to the bottom-up, no-computation and null families. We found overwhelming evidence (PXP = .886) that these computational quantities influenced top-down modulation, whereas the bottom-up (PXP = 0), no-computation (PXP = .113), and null (PXP = 0) hypotheses (fBOR = 0) were not validated. The probability that the model frequency in favor of top-down computational models was the same for all three groups in our sample was equal to $P(H_{F=} \mid y)$ = .968.

After showing the top-down controlled modulation of belief and PE+, we asked whether the aMFG and pMFG were differentially involved in these two distinct mechanisms. BMS revealed no clear evidence in favor of one family over the other (PXP = .343 and PXP = .657, fBOR = .677). Further between-group comparisons revealed that the probability that there were no underlying differences in model architecture was equal to $P(H_{F=} \mid y)$ = .828 when PTSD+ and PTSD- groups were compared, and $P(H_{F=} \mid y)$ = .796 when PTSD+ and nonexposed groups were compared.

Excessive belief suppression and alteration of reactive control in PTSD

To compare reactive and predictive control mechanisms between groups, we performed Bayesian model averaging (BMA) of the 14 models included in the computational top-down family for each group separately. This was possible because the DCM architecture that best explained our data was the same across all three groups. However, given that no differences were observed within the combined family, we summed the coupling parameters from aMFG and pMFG to reflect the coordinate action of the core control network. BMA provides both individual- and group-specific posterior distribution of coupling parameters, weighted for posterior evidence across all models in a family (see Methods).

Our main hypothesis was that individuals with PTSD prioritize belief of intrusive memories over online re-experiencing (PE+), to proactively suppress memory processing (i.e., imbalance hypothesis). A marker of suppression has been associated with more pronounced top-down negative coupling[11,25]. We therefore expected the imbalance in individuals with PTSD to be associated with more negative coupling during predictive versus reactive control. We computed the interaction between control (i.e., predictive vs. reactive) and group (PTSD+ vs. PTSD- or nonexposed). We found disproportionate negative coupling with the rHIP during predictive versus reactive control in the PTSD+ group, compared with the nonexposed group, $t(125) = -2.81$, $p_{\text{false discovery rate, FDR}} = .007$, posterior probability (Pp) = .999, 95% CI [-.99, -.17], and PTSD-, $t(99) = -2.17$, $p_{\text{FDR}} = .009$, Pp = .999, 95% CI [-.1.01, -.01]. We found a similar Control * Group interaction for the cHIP, when we compared PTSD+ with PTSD-, $t(99) = -3.23$, $p_{\text{FDR}} = .006$, Pp = 1, 95% CI [-1.20, -.29], and a trend toward significance when we compared PTSD+ with the nonexposed group, $t(125) = -1.62$, $p_{\text{FDR}} = .071$, Pp =.995, 95% CI [-.73, .07]. The same pattern emerged when we combined the two parts of the hippocampus (i.e., wHIP), with PTSD+ showing a greater imbalance between predictive and reactive control than the nonexposed, $t(125) = -2.49$, $p_{\text{FDR}} = .014$, Pp = .992, 95% CI [-.81, -.09], and PTSD-, $t(99) = -2.91$, $p_{\text{FDR}} = .007$, Pp = .998, 95% CI [-.1.06, -.19], groups. No differences were found in the PC when the PTSD+ group was compared with the PTSD-, $t(99) = -0.05$, $p_{\text{FDR}} = .477$, Pp = .540, 95% CI [-.48, .50], and nonexposed, $t(125) = 0.13$, $p_{\text{FDR}} = .477$, Pp = .586, 95% CI [-.42, .37], groups.

To further characterize these interactions, we explored the main effect of control and the simple effects of coupling parameters, running *t* tests for each group and each target region. Statistical details of these analyses are reported in Table 1, as well as in Fig. 4. In summary, we observed significant negative coupling during reactive control of the hippocampus in both the nonexposed and PTSD- groups, but not in the PTSD+ group. By contrast, predictive control over the hippocampus was observed in all three groups. When we compared predictive and reactive control within each group, we found significant higher inhibitory control of beliefs compared with PE, but only for the PTSD+ group in the rHIP, cHIP and wHIP. No differences were found in the other two groups (see Fig. 4 and Table 1). The PC was controlled proactively, but not reactively, in all three groups.

		DCM Coupling parameters														
			Reactive control *PE+*					Predictive control *Belief*					Predictive – reactive control				
		df	*t*	*p-fdr*	*Pp*	95% CI	BF	*t*	*p-fdr*	*Pp*	95% CI	BF	*t*	*p-fdr*	*Pp*	95% CI	BF
Nonexposed	**rHIP**	71	-2.88	**.012***	.999	[-.37 -.07]	9	-5.04	**<.001***	1	[-.46 -.21]	>1000	-0.98	.178	.884	[-.31 .10]	2.5
	cHIP	71	-1.84	.085	.968	[-.22 .01]	108	-4.02	**<.001***	1	[-.28 -.16]	>1000	-1.77	.068	.989	[-.41 .01]	7.4
	PC	71	-0.05	.478	.526	[-.15 .14]	1.76	-6.33	**<.001***	1	[-.83 -.44]	>1000	-4.51	**<.001***	1	[-.90 -.36]	>1000
	wHIP	71	-2.83	**.012***	.946	[-.28 -.05]	57	-5.58	**<.001***	1	[-.43 -.21]	>1000	-1.59	.076	.886	[-.33 .04]	6.6
PTSD-	**rHIP**	45	-0.63	.356	.764	[-.22 .12]	3.6	-1.78	**.041***	.99	[-.46 .05]	76	-1.09	.168	.943	[-.48 .15]	5.1
	cHIP	45	-3.44	**.007***	1	[-.62 -.17]	103	-2.01	**.027***	.99	[-.35 -.04]	6.4	1.66	.076	.977	[-.03 .47]	1.1
	PC	45	0.54	.356	.800	[-.14 .28]	2.8	-4.55	**<.001***	1	[-.88 -.37]	>1000	-3.62	**.001***	1	[-1.1 -.30]	9.1
	wHIP	45	-2.60	**.019***	.986	[-.39 -.06]	34	-2.21	**.019***	.98	[-.37 -.02]	97.5	0.18	.429	.557	[-.21 .27]	1.8
PTSD+	**rHIP**	54	0.76	.356	.861	[-.13 .30]	1.1	-5.44	**<.001***	1	[-.81 -.38]	>1000	-3.62	**.001***	1	[-1.1 -.32]	105
	cHIP	54	0.43	.362	.790	[-.16 .29]	3.6	-4.93	**<.001***	1	[-.66 -.29]	>1000	-2.93	**.004***	1	[-.89 -.18]	3.8
	PC	54	0.57	.356	.810	[-.13 .25]	1.15	-6.89	**<.001***	1	[-.76 -.43]	>1000	-4.54	**<.001***	1	[-.93 -.36]	>1000
	wHIP	54	0.66	.356	.750	[-.12 .28]	1.6	-6.06	**<.001***	1	[-.72 -.38]	>1000	-3.61	**.001***	.999	[-.94 -.30]	88

Table 1. Within-group BMA coupling parameter statistics. We first explored the negative modulation of PE+ (reactive control) and beliefs (predictive control) during top-down suppression. We then compared PE+ and belief coupling parameters within each group, to gauge the imbalance between reactive and predictive control (right part of the table). For both analyses, we also report *t* statistics, *p* values adjusted for false discovery rate correction (p_{FDR}; see Method), the group mean being different from zero (posterior probability, Pp; for the reactive vs. predictive control analyses, we report the probability of reactive and predictive control being significantly different)the bootstrapped 95% confidence intervals (CIs) of the mean, and the Bayes factor (BF). *Df*: degrees of freedom; rHIP: rostral hippocampus; cHIP: caudal hippocampus; PC: precuneus; wHIP: whole hippocampus.

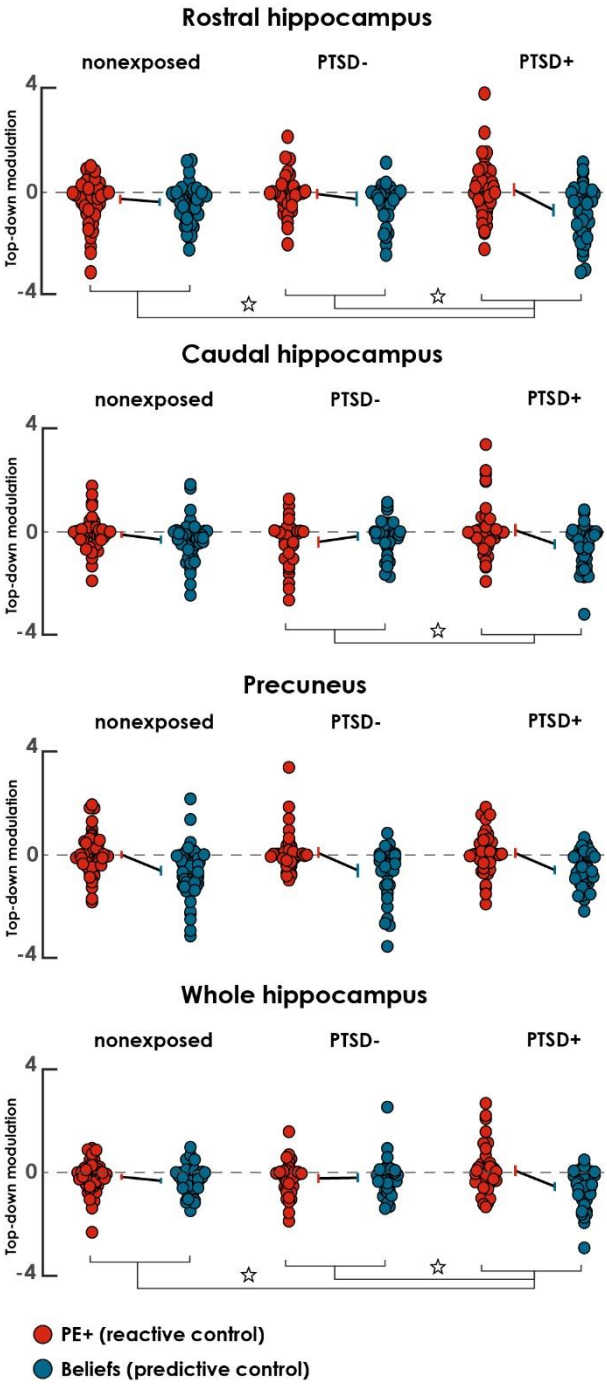


Figure 4. BMA of top-down coupling parameters during belief- and PE-driven suppression. Red and blue circles represent the modulation of the top-down coupling between the MFG and PE+ and belief target regions. Error bars represent the bootstrapped 95% CI of the group mean.

Excessive predictive control is related to re-experiencing and avoidance dimensions of PTSD but not transdiagnostic symptoms

We then examined whether the excessive of predictive control observed in individuals with PTSD could be specifically related to re-experiencing and avoidance symptoms, the two dimensions of PTSD presumably associated with such disruption, rather than to the general alteration of mental health. While intrusion and avoidance are two cardinal features of PTSD related to the traumatic memory, other symptoms associated with PTSD cross diagnostic boundaries. A recent study[26] examining trauma, anxiety and mood disorders found three transdiagnostic anxiety-related dimensions: anxious arousal, dysphoric arousal (i.e. tension), and general anxiety, and three transdiagnostic affect-related dimensions: anhedonia, mood and depression. We investigated the relationship between re-experiencing, avoidance, anxiety-related dimension, and affect-related dimension on one hand, and the imbalance of memory control mechanisms regulating the hippocampal activity on the other hand, in both the PTSD+ and PTSD- groups. We tested the hypotheses that excessive predictive control in the PTSD+ group was related to an increase in avoidance and intrusion, and that such negative relationship was significantly stronger than the relationship observed for anxiety- or affect-related dimensions, or the relationship observed in the same dimension but in the PTSD- group. Intrusion, avoidance, mood, anhedonia, dysphoric arousal, and anxious arousal symptoms were obtained from the PTSD checklist for DSM-5 (PCL-5)[27] and were adjusted for total symptom severity to ensure that the correlation with these dimensions were not confounded with PTSD severity. Depression and general anxiety dimensions were obtained using the Beck Depression Inventory and State Anxiety Inventory, respectively. After computing correlation between control imbalance in the wHIP and each of these symptoms, dysphoric arousal, anxious arousal, and general anxiety were summarized to reflect an anxiety-related dimension, while anhedonia, mood, and depression were summarized to reflect affect-related dimension.

In the PTSD+ group, we found that excessive predictive memory control significantly correlated with higher severity of avoidance ($R_{\text{-spearman}}$ = -0.32; 95% CI = [-.52 -.09]; *Z-val* = 2.27; p_{FDR} = .047) and marginally to intrusion symptoms after FDR correction ($R_{\text{-spearman}}$ = -0.26; 95% CI = [-.47 -.03]; *Z-val* = 1.84; p_{FDR} = .065). On the opposite, there was no significant relationship with the severity of both anxiety-related ($R_{\text{-spearman}}$ = 0.04; 95% CI = [-.08 .16]; *Z-val* = 0.55; p_{FDR} = .30) and affect-related ($R_{\text{-spearman}}$ = 0.09; 95% CI = [-.04 .23]; *Z-*

val = 1.09; p_{FDR} = .18) transdiagnostic symptoms (see Fig. 5). Crucially, we statistically compared the relationship that predictive control entertains with avoidance and intrusion in the PTSD+ group, to those entertain with trans-diagnostic symptoms (anxiety-related and affect-related dimensions). We used a boostrapping approach to obtain the confidence interval of the correlation difference and the p-value, respectively. Excessive predictive control was significantly more strongly related to re-experiencing symptoms than with anxiety-related (correlation difference 90% CI [-.62, -.14], *Z-val* = 3.09; p_{FDR} = .004) or affect-related (correlation difference 90% CI [-.49, -.04], *Z-val* = 2.36; p_{FDR} = .018) transdiagnostic clinical features. A similar pattern was observed for avoidance compared with anxiety-related (correlation difference 90% CI [-.66, -.22], *Z-val* = 3.6; p_{FDR} = .001) or affect-related (correlation difference 90% CI [-.52, -.13], *Z-val* = 2.82; p_{FDR} = .006) dimensions. Furthermore, excessive predictive control was significantly more strongly related to avoidance symptoms (correlation difference 90% CI [-.63, -.04], *Z-val* = 2.12; *p* = .034) in the PTSD+ than in the PTSD- group, although such difference in correlation between groups was not observed for re-experiencing symptoms (correlation difference 90% CI [-.51, .05], *Z-val* = 1.52; *p* = .13).

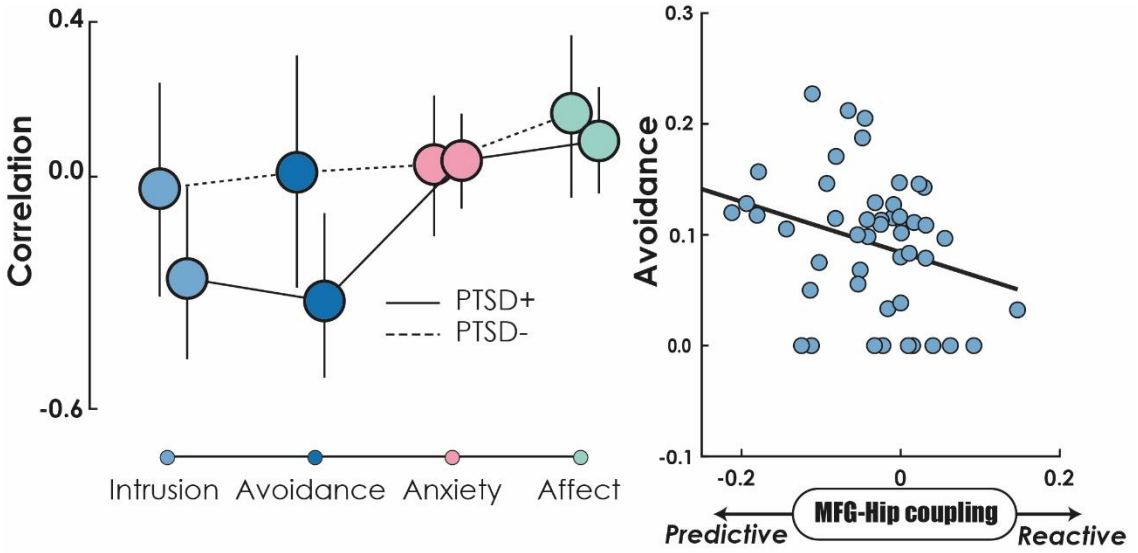


Figure 5. Left panel: correlations between the balance of memory control over the hippocampus (i.e., predictive – reactive control coupling parameters) and mental health features for PTSD+ (solid line) and PTSD- (dashed line). Error bars represents 95% boostrapped CI of the correlation (and thus indicate significance when they do not overlap with zero). Anxiety-related transdiagnostic features included anxious arousal, dysphoric arousal, and general anxiety, while affect-related transdiagnostic features included anhedonia, mood, and depression. Right panel: relationship between control imbalance in the whole hippocampus and avoidance symptom severity (adjusted for total symptom severity).

Imbalance between predictive and reactive control in PTSD reflects independent processes

Taken together, these findings suggest that individuals with PTSD cannot harmoniously balance predictive and reactive control in the hippocampus, unlike healthy individuals. This imbalance might reflect exaggerated predictive control applied in anticipation that prevents the deployment of reactive control. Contradicting this idea, however, predictive regulation of the hippocampus in PTSD+ was not related to reactive control ($R_{\text{-spearman}}$ = 0.01, 95% bootstrapped CI [-0.26, 0.30]).

Alternatively, despite serving the same down-regulation function of memory processes, predictive and reactive control can be conceptualized as two independent, yet downward forces, jointly mitigating hippocampal activity. These two directional forces can be projected on two distinct orthogonal axes (i.e., separated by a 90° angle) in a two-dimensional circular space (see Fig. 6, on the left). In this framework, the imbalance is reflected in the direction of the resultant vector combining the two forces. We fixed the 0° position at the bottom of the y-axis, and computed the direction of the resultant vector with respect to this optimally balanced position (see Methods). The angle of the resultant vector reflected an imbalance in favor of either predictive control (from 0° to 180°, moving anticlockwise) or reactive control (from 0° to -180°, moving clockwise).

In the hippocampus, we found a significant imbalance in favor of predictive control in the PTSD+ group (M = 33.35°; 95% CI [20.2°, 46.2°]) and the nonexposed group (M = 15.33°, 95% CI [4.55°, 26.51°]), but not in the PTSD- group (M = 6.86°; 95% CI [-9.17°, 23.8; see Fig. 6). When we compare the groups using circular statistics[28] (see Methods), we observed that the imbalance toward predictive control in the hippocampus increased significantly for PTSD+ compared with both the PTSD-, $t(99) = 2.10$, $p = .018$, 95% CI [-46.8°, -4.2°], and nonexposed, $t(125) = 1.74$, $p = .042$, 95% CI [-35.77°, -0.86°], groups. No differences were found between the nonexposed and PTSD- groups, $t(114) = 0.72$, $p = .235$, 95% CI = [-11.3°, 27.86°].

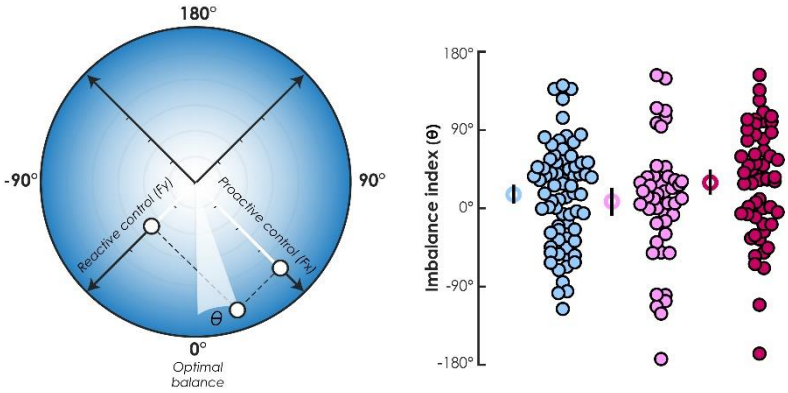


Figure 6. Left panel: circular projection of the resultant vector of predictive (i.e., belief) and reactive (i.e., PE+) control forces. 0° represents the optimum balance between these two orthogonal forces. The angle of the resultant force indicates imbalance toward either predictive ($\theta > 0°$) or reactive ($\theta < 0°$) control. The right panel shows the distribution of this imbalance index for each of the three groups. Error bars represents 95% boostrapped CI of the mean.

Discussion

To explain the persistence of intrusive traumatic memories and their avoidance, previous accounts of PTSD have largely focused on the disruption of memory functions[7,29]. More recently, analyses of brain connectivity analyses in individuals with PTSD during a memory suppression task revealed a lack of adaptive modulation of top-down control over memory processing in response to intrusive memory cues, suggesting that this persistence may additionally be rooted in the disruption of inhibitory control processes supporting active forgetting[10]. However, the origin of these control deficits remains unknown, and standard analyses of connectivity mask the hidden influence of predictive processing over control processes. Here, we suggest that abnormal predictive processing[6] constitute a unifying framework that links these two seemingly unrelated accounts of PTSD.

We showed that prediction of future memory control demand related to intrusions drives the flexible adaptation of memory suppression. These dynamic adjustments are orchestrated by a top-down inhibitory signal originating from the right DLPFC, which optimally balanced the suppression of the beliefs of future intrusive re-experiencing and their actual online emergence. This balancing is compromised in individuals with PTSD, but not in resilient or nonexposed individuals. We found that the disproportionate predictive inhibitory control over hippocampal activity based on beliefs, coupled with the reduction in reactive control based on PE+, was specifically related to cardinal features of PTSD related to the trauma, including

avoidance and traumatic re-experiencing. This finding echoes recent proposals suggesting that disturbances of predictive processing about threat are central to the expression of PTSD, including avoidance behaviors and traumatic re-experiencing[3,30]. Our findings shows that in PTSD, computations conferring higher value on predictions and beliefs than on outcomes also corrupt control processes, suggesting that maladaptive avoidance responses generalize to memory processes and nonthreatening situations.

Do these observations reflect a genuine, distinct deficit of reactive control in PTSD? The presence of a crossover interaction between control conditions and groups does not guarantee the existence of independent mental processes[31]. The disruption of reactive control may arise from exaggerated predictive control, and not reflect a genuine deficit in the online purging of intrusive memories. Extreme anticipation may prevent the control system from flexibly and adaptively adjusting its response when predictive attempts have failed, suggesting instead a single processing continuum between two modes[53]. This means that there may not necessarily be a second disrupted reactive control mechanism independent of predictive control. This hypothesis, however, seems unlikely, as we did not observe a negative relationship between the magnitudes of predictive and reactive control. Furthermore, we observed an imbalance after treating these two components of control as orthogonal yet downward forces originating from the same point of application (Fig. 6). This illustrates how a single control system could regulate two distinct computational quantities that are independently in the service of the same function (i.e., suppression of unwanted memories). However, these complementary processes take place within the same neurobiological system, which raises the question of how one (predictive control) may be enhanced (or at least preserved) when the other one (reactive control) is disrupted.

The ability to countermand the PE associated with intrusive memories may depend on the availability of executive control resources. Executive resources may be diminished in PTSD following gray-matter atrophy in the right DLFPC[34], or affected by disruption of the white-matter tracts originating from the prefrontal cortex[35]. PE increases attentional demand during learning in individuals with PTSD[4]. Thus, although limited executive functioning may allow for sustained predictive control in the background[33], it may proscribe the more demanding transient regulation of PE associated with intrusive memories. We did not observe any difference between the aMFG and pMFG with respect to predictive or reactive control, suggesting a general disruption of inhibitory executive functions. This finding fits

observations in the motor domain suggesting that both forms of control are coordinated and interact in the DLPFC[36].

Alternatively, the current findings may not reflect difficulties of the executive system, but alterations of the receptor system, which converts excitatory projections from the prefrontal cortex into local feedforward inhibition via GABAergic interneurons. It has been suggested that predictive and reactive forms of inhibitory control of the hippocampus are implemented via two distinct neuroanatomical pathways[14]. According to this model, predictive control processes may preferentially modulate the activity of rhinal inhibitory interneurons to gate inputs to the hippocampus, preventing the initiation of the retrieval process. The extent and nature of rhinal alterations in PTSD remain unclear, compared with alterations of the hippocampus proper. Reactive control, however, may activate CA1 inhibitory interneurons via the thalamic reuniens, a hippocampal subregion particularly involved in the regulation of pattern completion during memory retrieval[14]. Interestingly, studies conducted in rodents suggest that chronic stress affects GABAergic interneurons[37]. These are neurotransmitters that mediate memory control mechanisms in the hippocampus[38] and regulate the activity of dopamine PE neurons[39]. Alteration of this inhibitory function might therefore explain the excessive pattern completion and the lack of control over intrusive memories in individuals with PTSD.

We do not yet know whether the mechanisms identified here are related to the formation and persistence of traumatic memory traces in individuals with PTSD. Proactive avoidance of memories intrinsically implies the preservation of the related memory trace, maintaining the negative beliefs[40]. Furthermore, monitoring of the to-be-avoided representations increases paradoxical rebounds and the persistence of trauma-related memories[41]. Lastly, excessive interruption of hippocampal processing through predictive control may prompt the forgetting of safe contexts[42] associated with trauma reminders and contribute to the overgeneralization of fear. Previous TNT studies in healthy individuals have suggested that motivated forgetting is preferentially linked to the control of intrusive memories crossing the proactive gate[33]. Further investigations are required to evaluate whether the persistence of traumatic memory could be related to an inability to reactively countermand the neural activity associated with PE and involuntarily recall. On the one hand, PE increases the malleability of the memory trace[44] and its control might facilitate forgetting by promoting memory destabilization during the (re)consolidation mechanisms occurring during memory recall[45,46] . On the other hand, predictive coding models of the brain propose that memory recall arises from the disinhibition

of pyramidal cells encoding the bottom-up PE[47]. Such disinhibition is orchestrated by the hippocampus and its suppression might increase the plasticity of inhibitory engram and the silencing of neocortical traces[48].

Previous studies defined reactive control based solely on the presence of intrusive memories, without disentangling the confounding influence of predictive control dynamics, possibly leading to misinterpretations of the meaning of inhibitory control observed during memory intrusions. The absence of between-group differences with respect to the PC, previously associated with the suppression of intrusive memories in resilient individuals[10], further illustrates this point. Our neurocomputational approach overcomes this overlap and provides a partial answer to the longstanding question about the relationship between avoidance and memory suppression in PTSD.

Most of the recommended therapeutic treatments that have been shown to be effective for PTSD involve overcoming avoidance of the traumatic experience. Our findings suggest that this avoidance may result from the general disruption of hidden predictive operations engaged to infer and anticipate intrusive memories, biasing their control. Although our findings suggest that such bias is specifically related to trauma-related dimension of PTSD, and not to other transdiagnostic features related to affect or anxiety disorder, future studies would be needed to demonstrate the link between the development of a predictive control disorder and the development of the traumatic memory. Yet, this opens up possible new avenues for understanding the formation and maintenance of the traumatic engram in terms of predictive control disorder. New interventions designed to modulate and update the traumatic engram after it has been re-indexed in the hippocampus[29] should aim to restore the balance between predictive and error-driven control.

Materials and methods

Participants, materials and procedures

Detailed descriptions of participants, materials used, and task procedure can be found elsewhere[10]. The current study and analysis were performed on the same participants and data. We briefly describe the participants and the procedure here. The study was approved by the regional research ethics committee (Comité de Protection des Personnes Nord-Ouest III, sponsor ID: C16-13, RCB ID: 2016-A00661-50, clinicaltrials.gov registration number: NCT02810197). All participants gave their written informed consent before taking part. The study lasted from 13 June 2016 to 7 June 2017. The exposed groups did not differ on the length of time between the date of the Paris attacks and the date of inclusion in the study (PTSD- = 1.14 ± 0.18 years, PTSD+ = 1.18 ± 0.22 years). Participants were aged 18-60 years, they were all right-handed and spoke French. None of them reported any prior psychiatric (e.g., psychotic, bipolar, or obsessive-compulsive disorder) or neurological diseases, traumatic brain injury (with loss of consciousness > 1 hr), alcohol or substance abuse (other than nicotine), or MRI contra-indications. In addition to the above-mentioned criteria, both non-exposed (N = 73) and exposed (N = 102) participants were not included if they had any history of PTSD, depression or anxiety disorder prior to the attacks. A medical doctor screened participants for the inclusion/exclusion criteria during a medical examination. Among the exposed participants, 55 has been diagnosed with PTSD (in its full or partial form[49]) and 47 had not (they met DSM-5 Criterion A, indicating that they had experienced a traumatic event, but did not present any re-experiencing symptom or experience functional impairment). Partial forms must include re-experiencing symptoms (Criterion B), with persistence of the symptoms for more than 1 month (Criterion F), causing significant distress and functional impairment (Criterion G). The diagnosis was performed using the Structured Clinical Interview for DSM-5 (SCID)[50], conducted by a trained psychologist and supervised by a psychiatrist. Severity of PTSD symptom clusters was assessed with the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5)[51]. Severity of depressive and general anxiety symptoms was quantified using the Beck Depression Inventory and State Anxiety Inventory, respectively. Further demographic and clinical information can be found here[10].

Before fMRI acquisition, participants intensively learned neutral French word-object pairs. This overtraining procedure was intended to ensure that the cue word would automatically trigger the retrieval of the associated object. We recorded fMRI activity during the TNT task. During this task, 36 cue words repeated 8 times were displayed either in green (think condition) or red (no-think condition). During think trials, participants had to visualize and recall the associated object with as many details as possible. During no-think trials, participants had to try and prevent the memory of the object from entering awareness and maintain their attention on the cue word. If the object came to mind anyway during suppression attempts, they were asked to push it out of their mind and to report at the end of the trial that the reminder had elicited awareness of its paired object, allowing us to pinpoint which no-think trials triggered intrusions. It should be noted that before the TNT trials, we tested memory for word-object pairs and discarded any forgotten pairs from subsequent analyses. The data of two participants (one nonexposed and one PTSD-) were excluded from the final analyses, as they had an unusually low number of remaining pairs, making it impossible for us to calculate an item-specific belief computational model (see below).

Computational Modeling

We used computational modeling to investigate participants' beliefs about upcoming intrusive memories in the no-think condition of the TNT task. Taking the *observing the observer* meta-Bayesian approach[17] one step further, our aim was to *observe the observer observing him- or herself*. According to this approach, agents use a perceptual model to make inferences about the hidden states that control the world. The observation (or response) model describes the relationship between inferred hidden states and behavioral outcomes. In our models, inputs *u* and outcomes *y* were binary:

$$u^{(t)} \in \{0,1\}; \;\; y^{(t)} \in \{0,1\} \qquad \text{Eq. 1}$$

where 0 corresponds to nonintrusion and 1 to intrusion at time *t*. As our aim was to model participants' beliefs about their own intrusion ratings during the TNT, input *u* at time *t* was outcome *y* at time *t-1:*

$$u^{(t)} = y^{(t-1)} \qquad \text{Eq. 2}$$

To model individual time series of internal beliefs, we used the HGF and RW models implemented in the TAPAS toolbox (available at https://www.tnu.ethz.ch/de/software/tapas.html), which applies variational Bayesian inversion to infer hidden states maximizing the log-model evidence (LME).

Perceptual Models

Two-level hierarchical Gaussian filter: we used a two-level HGF as a perceptual model. Developed by Mathys et al.[21], the HGF assumes that agents form internal beliefs in a hierarchical fashion. Implementing a variational approximation approach, the HGF allowed us to estimate trial-by-trial trajectories of internal beliefs at multiple levels. The lowest level corresponds to participants' beliefs about whether they were experiencing a memory intrusion or not (x_1). As $u^{(t)}$and $y^{(t)}$ are binomial, x_1 assumes a Bernoulli distribution. Accordingly, first-level beliefs x_1 are the logistic sigmoid transformations of second-level beliefs x_2 which, by contrast, are unbounded:

$$x_1^{(t)} \sim Bernoulli\left(\frac{1}{(1+exp(-x_2^{(t)}))}\right) \quad \text{Eq. 3}$$

The second level (x_2) corresponds to participants' internal beliefs about the volatility of memory intrusions experienced during the TNT task: x_2 is denoted as a Gaussian random walk whose step size is controlled by the free parameter ω. The resulting beliefs assume Gaussian distributions described by their sufficient statistics: posterior mean μ and uncertainty σ (i.e., variance):

$$x_2^{(t)} \sim N(x_2^{(t-1)}, exp(\omega)) \quad \text{Eq. 4}$$

where the ω parameter controls the variance of x_2, shaping the magnitude at which beliefs are updated. We used the superscript ^ to indicate prior internal beliefs. For example, $\mu^{(t)}$ represents posterior internal beliefs at Trial t, and $\hat{\mu}^{(t)}$ represents internal beliefs prior to the outcome $y^{(t)}$ (intrusion or nonintrusion).

The variational approximation underlying the HGF model fitting allowed participant-specific free parameters to be estimated, along with the trial-by-trial trajectories of internal belief updating, which were determined by the participants' sets of parameters. Crucially, the

updating of second-level beliefs $\mu_2^{(t+1)} - \mu_2^{(t)}$ in the model is proportional to ascending first-level prediction errors weighted by their uncertainty:

$$\mu_2^{(t+1)} - \mu_2^{(t)} \propto \Psi^{(t)}\delta^{(t)} \quad \text{Eq. 5}$$

where Ψ is a weighting factor representing the inverse of second-level belief precision $\pi_2^{(t)}$ (i.e., uncertainty):

$$\Psi^{(t)} = \frac{1}{\pi_2^{(t)}} \quad \text{Eq. 6}$$

This quantity is modulated by the ω parameter, and δ represents PE, namely the difference between beliefs after and before presentation of a stimulus:

$$\delta^{(t)} = \mu_1^{(t)} - \mu_1^{(t-1)} \quad \text{Eq. 7}$$

As participants were instructed to report whether or not they experienced a memory intrusion at time *t*, posterior beliefs are equal to the outcome:

$$\mu_1^{(t)} = y^{(t)}$$

The belief updating equation allowed us to estimate participants' predictions $\hat{\mu}^{(t)}$ about the outcome $y^{(t)}$ before it occurred. Importantly, as $\mu^{(t-1)}$ corresponds to prior internal beliefs about the outcome (i.e., sigmoid transformation of μ_2),

$$\hat{\mu}_1^{(t)} = \frac{1}{1 + exp(-\mu_2^{(t-1)})} \quad \text{Eq. 8}$$

PE (or $\delta^{(t)}$) represents the divergence between the real outcome (i.e., intrusion/nonintrusion) and the predicted one:

$$PE^{(t)} = y^{(t)} - \hat{\mu}_1^{(t)} \quad \text{Eq. 9}$$

Next, according to Eq. 5, the updating of posterior beliefs about the tendency to experience intrusions (i.e., $\mu_2^{(t)}$) is driven by the quantification of prediction failure (i.e., $PE^{(t)}$), weighted

by uncertainty about the beliefs (i.e., $\frac{1}{\pi_2^{(t)}}$ in Eq. 6). Thus, when beliefs are more uncertain, PE has a greater impact on belief updating, improving future predictions about upcoming trials. Importantly, by shaping the uncertainty of beliefs, ω plays a crucial role in their updating.

For model fitting, we used prior parameters defined in de Berker et al.[52], who conferred high variance on ω priors (mean: -3, variance: 16) in order to efficiently catch any possible between-participants variability on this parameter. It should be noted that 10 of the 173 participants included in this study showed no modification of the ω parameter in its prior state (i.e., -3; see Fig. 2e). This absence of departure from the prior mean was due to the presence of a stochastic occurrence of intrusion rating (with a mean consistently close to .5 throughout the task), prohibiting the consistent updating of this parameter. It should, however, be noted that the belief trajectories were still valid for these participants and could be used to infer model accuracy or in subsequent connectivity analyses.

Kalman filter: to include the hypothesis that internal beliefs about experiencing intrusions are uncertain, dynamically updated, but nonvolatile (contrary to HGF), we included a KF[20,53] in our model space. Like the HGF, the KF estimates the trial-by-trial weighting of PE in belief updating, but in this model, beliefs are not hierarchical, and uncertainty therefore remains constant during learning. In the KF framework, beliefs about experiencing an intrusion $\hat{\mu}$ are updated as follows:

$$\hat{\mu}^t = \hat{\mu}^{(t-1)} + K\delta^{(t-1)} \quad \text{Eq. 10}$$

where K is the Kalman gain, representing trial-by-trial learning. The gain is modulated by two free parameters (π and ω) that encode belief reliability and uncertainty:

$$K^t = \frac{K^{(t-1)} + \pi\omega}{K^{(t-1)} + \pi\omega + 1} \quad \text{Eq. 11}$$

These two free parameters model two different aspects of belief updating: π quantifies how far beliefs can be trusted, based on previous trial history, and ω quantifies the process variance (i.e., how uncertain the beliefs are).

Rescorla-Wagner: to include the hypothesis about the role of trial-by-trial weighting of PE during intrusion control, we compared the HGF2 and KF models with a traditional

reinforcement learning model: the RW[19]. Briefly, RW, HGF and KF share a similar general update equation[21], defined by a weighting factor and prediction error. However, RW assumes a participant-specific fixed learning rate α:

$$V^{(t)} = V^{(t-1)} + \alpha(\lambda - V^{(t-1)}) \quad \text{Eq. 12}$$

where V is the prediction and $(\lambda - V^{(t-1)})$ the prediction error (i.e., divergence between real outcome λ and prediction at previous trial).

Source Models

Perceptual models were built using intrusion ratings either from the entire sequence of trials (state model), or separately for each pair of word-object memories (item model), including eight repetitions in total. After concatenation of item trajectories, state and item belief trajectories were linked to an observation model either separately or in combination. Observation models linked the inferred hidden states to the outcomes, describing the probability of observing an outcome y given model parameters. For each model in the perceptual model space, we built an observation model based on beta density probability distributions:

$$p(y|\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{(\alpha-1)}(1-y)^{(\beta-1)} \quad \text{Eq. 13}$$

where θ refers to participants' beliefs estimated through the different perceptual models, Γexpresses a Gamma function, $\alpha = \theta * \nu$, and $\beta = \nu - \alpha, \nu$ is a participant-specific free parameter (i.e., inverse decision noise regulating beta density width, estimated during model fit). Here, the observation model described the accuracy of internal beliefs about outcomes (i.e., intrusions). Note here that the beta observation model performed better than other observation function such as the softmax response model (because the log-probability of choice of the beta observation model does not change sharply around belief equal to 0.5, preserving model accuracy). However, although our data do not involve such extreme cases, this model contains the slight absurdity that when beliefs approach certainty (i.e. near 1 or 0), the corresponding probability of choice starts to sink again. For all three models in the perceptual model space (i.e. RW, KF, and HGF), we built the following three source models.

- The *state* source model hypothesized that belief θ ($\hat{\mu}_{1s}$ for HGF, $\hat{\mu}_s$ for KF and V_s for RW) at trial *t* was influenced by previous trial history, irrespective of the content of the specific item.

- For the *item* source model, we extracted beliefs for each specific no-think item. Throughout the TNT, up to 18 different items (i.e., object-word pairs) were repeated (on average, 16.29 ± 2.18, no group differences, after accounting for error or absence of criterion recall test before trial phase). For each item *i*, we estimated the trajectories of participants' predictions based exclusively on the item's specific history. For these item-specific models, *t* in Eq. 1-15 refers to the number of times the item *i* was repeated, instead of the overall no-think trial count. The trajectory of item-based beliefs is referred to as $\hat{\mu}_{1i}$ for HGF, $\hat{\mu}_i$ for KF, and V_i for RW. After estimations, these separated item-based beliefs were concatenated to form a single trajectory.

- In the *combined* source model, we considered a scenario in which participants combined state and item beliefs to improve prediction accuracy. A joint posterior distribution with mean $\hat{\mu}_c$ was created (starting from the second repetition of each item) by summing the two types of beliefs, weighted for their respective accuracy, and dividing the result by the sum of the variances:

$$\theta = \hat{\mu}_c = \frac{\hat{\mu}_{1s}\hat{\pi}_{1s} + \hat{\mu}_{1i}\hat{\pi}_{1i}}{\hat{\pi}_{1s} + \hat{\pi}_{1i}} \quad \text{Eq. 14}$$

  This combined model hypothesized that participants lent more weight to the most accurate (i.e., least uncertain) source of beliefs when creating combined beliefs $\hat{\mu}_c$. For the KF and RW models, we averaged $\hat{\mu}_s$ and $\hat{\mu}_i$, and V_s and V_i, respectively.

Model estimation and accuracy

The final model space therefore included nine models: state-HGF2, item-HGF2, combined-HGF2, state-KF, item-KF, combined-KF, state-RW, item-RW, and combined-RW. Free perceptual parameters and corresponding belief trajectories were estimated using a quasi-Newtonian optimization algorithm[21]. For state, item, combined trajectories of belief, we computed model accuracy using the sum of the negative log-likelihood of the choice probability.

Validation of computational modeling

Model falsification

A common issue in computational modeling is how to assess the performance of a set of different models in generating plausible data, given that generative and predictive performances of a model can sometimes be dramatically different[22]. This is an important step that allows the models that best generate plausible data to be identified and those with poor generative performances to be rejected. This procedure is known as *model falsification*[22].

The goal of these simulations was to establish whether the models were able to generate the behavioral reduction in intrusion proportion that we observed across the four blocks of the TNT task (see Fig. 2). We designed a virtual experimental setting with 144 suppression cues distributed across 4 TNT sessions, as in our real experiment. We started with a belief of .5 for the first trial, and at each new simulated trial, we generated a new belief based on the perceptual model considered and randomly drawn corresponding perceptual parameters. A suppression parameter was introduced to simulate memory suppression and to avoid the tilting of belief trajectories toward 1. This parameter was initially fine-tuned separately for each model using a grid search to minimize the difference between simulated data and real intrusion profile. After applying this suppression factor to the generated belief, and adding some noise, we computed the negative log-model accuracy of previous responses using the beta observation model (i.e. summing all trials response log-probabilities up to the new one), and generate a new response (i.e. *yes* or *no*) depending on log-model accuracy improvement (i.e. we selected the response for this new trial that best improved the overall log-model accuracy). The inverse decision noise parameter (ν; see above) of the beta observation model was fixed to e^0 (i.e., 1), allowing the mapping to be unbiased toward a preferred outcome.

We simulated 200 virtual participants using this procedure, and repeated the virtual experiment 100 times using perceptual parameter randomly drawn from a Gaussian priors distribution tailored to match our own data (to sample plausible parameters), resulting in 20000 simulations for each of the nine computational models. Then, binary rating generated for each of these 200 simulations were averaged across repeated sampling and summarized as intrusion proportion across the 4 artificial TNT sessions. We tested the relationship with the real intrusion proportions for our cohort by computing both the mean difference (MD) and the mean correlation (MC) between real and simulated intrusion ratings across the 200 virtual participants. While the MD between simulated and fitted parameters is informative of the

absolute distance between real and simulated intrusion ratings, MC indicates whether simulated intrusions mimick the decrease in intrusion rating across normally observed across TNT sessions. We found that for HGF2, both state (MD = .069 ± .02; MC = .543 ± .05) item (MD = -.008 ± .01; MC = .367 ± .03) and combined (MD =.054 ± .02; MC = .575 ± .05) models were able to generate data both intercepting the session-wise mean intrusion rating and mimicking the decrease in intrusion proportion across the TNT blocks. Concerning the RW models, only the item model generated the expected patterns of intrusions (MD = -.004 ± 0.01; MC = .769 ± .03). While both state and combined models simulated intrusions showed acceptable correlations with real intrusion data (state: MC = .765 ± .03; combined: MC = .274 ± .05), both failed in intercepting the session-wise mean of the real intrusion data (combined: MD = .183 ± .02; combined: MD = .140 ± .01). Similarly to RW, also for KF only the item model generated the expected patterns of intrusions (MD = -.019 ± 0.01; MC = .769 ± .03), while both state and combined models showed acceptable correlations (state: MC = 889. ± .01; combined: MC = .292 ± .01, $p<.001$) but not mean differences (state: MD = .292 ± .01, combined: MD = .278 ± .01). The main outcomes of this model falsification analysis can be found in Fig. 2.

Recovery analyses

Given this evidence for the generative performances of our models, we addressed another possible pitfall in the model selection workflow: the ability of a set of models to recover their trajectories of belief and the associated perceptual parameters. This analysis further tests the generative performance of a model, by verifying whether the fitting procedure produces meaningful trajectories and/or parameters, namely the true parameters and the corresponding trajectories used to generate the data[23]. We fitted the different models to the synthetic data, in order estimate the trajectories and the free parameters.

For trajectory recovery, we computed the correlation between fitted and simulated trajectories. We then identified the fitted model among competitors that has the maximum correlation with the simulated trajectory (coding 1 for the best model, and 0 otherwise), and averaged these outcomes across simulations. We computed the inversion matrix, to ensure that the belief fitted by a given model was the model that most likely has generated those beliefs (i.e. reverse inference; Fig. 2).

When comparing computational models, it is also important to verify the reliability of the model selection criterion for identifying the true generative model within a set of competitive models, and ensure that this selection is not biased in favor of one particular model[22,23]. This procedure, known as model recovery, consists in simulating data with one specific model and then comparing the predictive performances (i.e. model accuracy) of a set of different models using Bayesian inference. For each of the 200 virtual participants, we first summed the model accuracy across the 100 random sampling. We then identified, for each simulated model, the best fitting model associated with the maximum accuracy, and summarized the probability into a confusion matrix to create the corresponding inversion matrix[23] (see Fig. 2).

For parameter recovery, we computed the correlation between simulated parameter that generated the data, and the corresponding fitted parameters. This correlation was computed for each of the 200 virtual participants, using 100 randomly sampled free parameters (see above), and then averaged across virtual participants. We found that HGF models had the best overall ability to recover the parameter ω, with small correlations for the state model ($r(98)$ = .263 ± .08, p = .008) and moderate correlations for the item ($r(98)$ = .395 ± .08, p < .001) model. Significant recovery of α was found in RW models for the state ($r(98)$ = .268 ± .13, p = .007), but not for item ($r(98)$= .013 ± .10, p = .898) model. No significant correlations were found between simulated and fitted ω (state: $r(98)$ = .008 ± .10, p = .937; item: $r(98)$ = .001 ± .09, p = .992) and π (state: $r(98)$ = -.004 ± .09, p = .968; item: $r(98)$ = -.007 ± .09, p = .945) in KF models (see Fig. 2).

Computational Dynamic Causal Modeling

Regions of interest

Details about fMRI acquisition and preprocessing can be found in Mary et al.[10]. DCM entails a priori selection of regions of interest (ROIs). There is evidence for a central role of the right PFC, particularly the MFG, in inhibiting the memory system during motivated forgetting[11,25,43]. The ROIs included in the DCM models were aMFG and pMFG, rHIP and cHIP, and PC. We initially selected the ROIs from the Brainnetome atlas (BNA[54], http://atlas.brainnetome.org/), which is a fine-grained connectivity-based atlas featuring 210 cortical and 36 subcortical cross-validated brain regions, defined in Montreal Neurological Institute (MNI) space. The aMFG region included A46 (center coordinates: x =

28, y = 55, z = 17) and A9/46v (x = 42, y = 44, z = 14), pMFG included A9/46d (x = 30, y = 37, z = 36) and A8vl (x = 42, y = 27, z = 39), rHIP and cHIP corresponded to two ROIs (x = 22, y = -12, z = -20 and x = 29, y = -27, z = -10), and PC corresponded to dmPOS (x = 16, y = -64, z = 25). The MNI coordinates of the five ROIs were projected onto participants' native space using the deformation field, without any spatial warping or smoothing of the functional images, to ensure maximum accuracy. However, for there to be sufficient demarcation between the aMFG and mMFG signals, aMFG coordinates were initially limited to y > 35 mm, and pMFG coordinates to y < 25 mm.

For the DCM analysis, we summarized the signals for each participant and each of these ROIs from the averaged time series of 30 contiguous voxels (1012.5 mm3) that were the most significantly related to the main task around the maximum activation peak (using no-think > think contrast for aMFG and pMFG, and no-think < think contrast for memory regions)[25]. To this end, a univariate analysis was conducted on the timecourse of each native space ROI for each participant, by implementing a general linear model (GLM) in SPM12. The voxelwise fMRI time series were high-pass filtered, with a cut-off period of 128 s. Task-related regressors were created by convolving a box-car function at the onset of cue words with the canonical hemodynamic response function. Further regressors of no interest included the six realignment parameters to account for motion artefacts, session dummy regressors, and filler item regressors (i.e., no button press, or no recall during the final criterion test or during think trials). fMRI time series autocorrelations were corrected by entering a first-order autoregressive model of temporal autocorrelation of noise and a white-noise model was estimated using restricted maximum likelihood. The data were then adjusted for confounds, filtered, and whitened using the estimated temporal autocorrelation of noise to correct for non-sphericity. Beta parameters for think and no-think conditions were estimated during a second pass of the general linear model with the ordinary least square method, and used to calculate participant-specific *t* maps for each ROI.

Neural and hemodynamic models for DCM

DCM[73] allows changes in effective connectivity between a set of brain regions to be inferred by creating and comparing different hypothesis-driven generative models of neural dynamics. It relies on the following general bilinear state equation for these dynamics:

$$\frac{dx}{dt} = \left(A + \sum_{j=1}^{m} u_j B^{(j)} \right) x + Cu \qquad \text{Eq. 15}$$

Given *m* known inputs, the hidden neural dynamics ($\frac{dx}{dt}$) are estimated by relating the activity of each region to the activity of other regions, via 1) intrinsic connections in the absence of experimental manipulations (A matrix), 2) j[th] modulatory input u_j operating on intrinsic connections during experimental conditions (B matrix), and 3) extrinsic input driving activity in the network (C matrix). These neural models are then combined with a hemodynamic model describing the mapping of neural activity onto the BOLD response observed during fMRI (i.e., the Balloon model[55]). The neural and hemodynamic model parameters are estimated through variational Bayes under Laplace approximation, which optimizes model evidence by minimizing free energy and ensures Gaussian posteriors[56].

Two modularity input functions operated on intrinsic connections. The first one corresponded to a boxcar function encoding no-think trials onset and duration, and whose height was parametrically modulated by internal beliefs ($\hat{\mu}_c$, see Eq. 14). The second corresponded to a boxcar function reflecting only intrusive trials, parametrically modulated by PE (see Eq. 9). This allowed us to investigate how the discrepancies between internal beliefs and intrusive outcomes were reactively processed by the memory control system, our primary interest. PE was therefore only positive here (PE+), meaning that negative and positive coupling parameters could be interpreted as such. It should, however, be noted that the extent and the sign of the posterior coupling parameters were estimated with respect to the implicit baseline (i.e., unmodeled signal). Here, the neural dynamics were only modeled during no-think trials. The implicit baseline included think trials, and the coupling parameters therefore reflected the modulation of coupling with respect to a baseline corresponding to a mixture of rest (i.e., no stimulation) and memory retrieval. Given that our design included few resting periods, this procedure ensured better isolation of inhibitory mechanisms during memory control. The parametric modulators were not orthogonalized, and were extracted from the winning computational model (i.e., combined-HGF2).

DCM model space

All the models assumed bidirectional intrinsic connections between all five regions in the A matrix. This was confirmed by a preliminary analysis that only modeled driving inputs[57]. We created 42 DCM models, which could be divided into three families of fourteen models each and a null family containing two models. The first family (computational top-down modulation family; Fig. 3A, top left) hypothesized that the modulation of PE+ and beliefs occurs during top-down coupling originating from the source regions (i.e., aMFG and pMFG) and targeting memory regions (i.e., rHIP, cHIP and PC). This family could be further divided into two subfamilies encoding different hypotheses on the involvement of aMFG and pMFG in either predictive or reactive control. More specifically, the first subfamily contained seven models encoding all the possible pathways from MFGs to target regions (Fig. 3B), hypothesizing that aMFG and pMFG are involved in reactive (i.e., PE+ modulation) and predictive (i.e., beliefs modulation) control, respectively. Importantly, while beliefs were computed before the actual outcome, including therefore in the no-think trials, PE+ only occurred when participants experienced an intrusion. For this reason, in the first subfamily, intrusion inputs entered the aMFG, while no-think inputs entered the pMFG. The opposite scenario was hypothesized in the second subfamily, with the aMFG and pMFG receiving inputs from no-think and intrusion cues, and modulating control of belief and PE+, respectively. The second family (computational bottom-up modulation family; Fig. 3A, right panel) hypothesized that computations modulate the bottom-up connections from target to source regions, with analogous subdivisions regarding the involvement of aMFG and pMFG with respect to belief and PE+ modulation. The third family (no-computation modulation family; Fig. 3A, bottom left panel) contained 14 models including modularity input functions with no further parametric modulation. Finally, a fourth family containing two null models was added to verify the hypotheses that our modulatory parameters did have an impact on connections, compared with models that did not include these additional modulations (Fig. 3A, bottom right).

Bayesian model selection and averaging

BMS compares different generative models, in order to select the most probable one. This allows competitive hypotheses on the hidden mechanisms that generated the data[58] to be tested. Here, for both computational and DCM model comparisons, we used a random-effect

BMS (i.e., assuming that models can differ between participants) and a free energy approximation of the LME, accounting for both the accuracy and complexity of the models[58]. Interestingly, BMS can be used to compare different families of models, where the model space is partitioned into several models sharing some common underlying hypotheses. For DCM BMS analyses, we first computed the log-family evidence, which summarizes the evidence for models belonging to a given family, assuming prior and posterior additivity of model probabilities into family probabilities, as well as a uniform prior within families[59]. We then compared this evidence using random-effect BMS implemented in the VBA toolbox[60]. Besides computing the probability of one model being more likely than the others in the model space (i.e., exceedance probability, XP), the VBA toolbox estimates the probability that potential differences in model frequencies are due to chance (i.e., BOR, BOR). XP and BOR can then be used to compute the PXP, which quantifies the probability of one model being more frequent than others in the model space, above and beyond chance[58].

Despite the remarkably high PXP for the whole sample, BMS did not guarantee that the same model was uniformly the best in all three groups. Traditionally, independent RFX-BMS has been used to establish the winning model in each separate group. However, this approach does not test the hypothesis that the same model optimally describes data in the different groups. To test this hypothesis, we adopted a recent method[58] implemented in the VBA toolbox[60], which allows between-group model comparisons. This technique computes the probability that different groups are sampled from a single population in which the elected model best explains the data.

We performed BMA across the 14 DCM models that belonged to the wining computational top-down modulation family (see Results section). BMA yields posterior coupling parameters specific to each participant that are weighted by participant-specific posteriors. The optimum model within the selected family is treated as a random effect across participants[61]. For each participant *s* belonging to the group g (i.e., nonexposed, PTSD-, or PTSD+), the averaged parameters across the 14 models of the family, $P(\theta_{s\in g} \mid Y, m \in f_D)$, are computed by weighting the participant's posteriors for each model m in the family (i.e., $P(\theta_s \mid y_s, m)$ by the posterior probabilities that participant *s* uses model *m* (i.e., $P(m_s \mid Y_g)$:

$$P\left(\theta_{s\in g} \mid Y_g, m \in f_D\right) = \sum_{m\in f_D} P\left(\theta_{s\in g} \mid y_{s\in g}, m\right) P\left(m_{s\in g} \mid Y_g\right)$$

Eq. 16

where Y_g is the dataset of the whole group g, containing data for each participant in the group, $y_{s \in g}$. Importantly, a separate analysis was performed for each group, to ensure that the participant's posterior probabilities $P(m_s \mid Y_g)$ were derived from his or her group's distribution. It should be noted that this was possible because the computational top-down modulation family outperformed the other families in all three groups, and no statistical differences were detected between groups with respect to the preferred model architecture (see Results section). Statistical analyses were performed on BMA coupling parameters using one-tailed *t* tests according to a priori hypotheses, in the three target memory regions (rHIP, cHIP, PC), as well as the wHIP (i.e., four regions in total). Four effects were tested:

1) Control * Group interactions comparing the control effect (predictive - reactive) in PTSD+ with both PTSD- and nonexposed in all four regions (i.e., 8 tests in total):

2) Control effect (predictive - reactive) in all three groups and four regions (i.e., 12 tests in total);

3) Reactive negative coupling in all three groups and four regions (i.e., 12 tests in total);

4) Predictive negative coupling in all three groups and four regions (i.e., 12 tests in total).

To control for Type I error across multiple tests, *p* values were adjusted for each of these effects, using FDR correction. For completeness, we also computed the Pp of the groups' coupling parameters, as well as the bootstrapped 95% CI of the mean. In addition, we also report Bayes factors (BF) as effect size in Table 1, using a Markov chain Monte Carlo (MCMC) method[62]. BF represent the likelihood of suppression effects for each within-group comparison. Based on this hypothesis, we defined a region of practical equivalence (ROPE) set as a Cohen's d effect size greater than "0.1". The MCMC method generated 90,000 credible parameter combinations that are representative of the posterior distribution. Then, the BF was estimated as the ratio of the proportion of the posterior within the ROPE relative to the proportion of the prior within the ROPE. The conventional interpretation of the magnitude of the BF is that there is substantial evidence for the alternative hypothesis when the BF

ranges from 3 to 10, strong evidence between 10 and 30, very strong evidence between 30 and 100, and decisive evidence above 100.

Imbalance analysis

We projected neurocomputational markers of predictive and reactive control of intrusive memories onto two orthogonal axes of a polar coordinate system (see Fig. 4B). Angular coordinates were expressed in degrees between -180° and +180°) with a 0° reference point at the bottom of the y-axis (i.e., 0° to 180° anticlockwise and 0° to -180° clockwise). The first axis (+45° to -135°) represented predictive control (PC). Negative PC coupling values were projected on the +45° direction, and positive PC coupling parameters onto the opposite -135° direction. The second axis (+135° to -45°) represented reactive control (RC). Negative RC coupling values were projected onto the -45° direction, and positive RC coupling parameters onto the opposite +135° direction.

For each participant, we calculated the resultant force (RF) combining predictive and reactive forces. The RF represents the vector sum of a set of forces. Given two forces F_{PC} and F_{RC}, characterized by known angles α_1 and α_2 from 0° on the y-axis of a circle and the x and y Cartesian components (Fx_{PC}, Fx_{RC} and Fy_{PC}, Fy_{RC}), the RF's Cartesian components can be obtained as follows:

$$F_{Rx} = Fx_{PC} + Fx_{RC};$$

$$F_{Ry} = Fy_{PC} + Fy_{RC}; \qquad \text{Eq. 17}$$

In our analyses, we focused on the RF's direction, not its magnitude. The RF represents the summative effect of predictive and reactive vectors of force. As the two forces were applied in different directions, yet both pointing downward, the 0° position represented the equilibrium point. The more the RF approached the 0° direction, the more balance the two forces were. To obtain an imbalance angle (IB) for each participant, we computed the angle θ between the RF and the 0° position using trigonometry:

$$\mathrm{IB} = \theta = \tan^{-1}\left(\frac{F_{Ry}}{\mathrm{F_{Rx}}}\right) \qquad \text{Eq.20}$$

Interestingly, both predictive and reactive negative coupling parameters reflected downward, yet orthogonal forces, originating from the same point of application. This illustrates how a unique control system may suppress memory processing according to two independent but complementary processes serving the same function.

A common issue in circular statistics is the arbitrary choice of the 0° position and the sense of rotation, which can lead to misleading conclusions when dealing with multiple measures. The mean angle θ cannot be computed from the arithmetic mean of all sampled angles. We used the Circular Statistics toolbox in MATLAB[63] to compute the mean angle θ across participants in each group. Confidence intervals were also computed by bootstrapping the estimation of this group mean 2000 times. Group comparisons were performed using Watson's two-sample tests, a nonparametric version of the two-sample *t* test for circular data. For all group comparisons, alpha was set at .05.

Data availability

All the raw behavioral and imaging data are archived at the GIP Cyceron center in Caen and are part of an ongoing longitudinal research project.

Code availability

Computational models were implemented in the TAPAS toolbox (https://www.tnu.ethz.ch/de/software/tapas.html). Preprocessing of fMRI data and first-level DCM analysis were performed with SPM12 (https://www.fil.ion.ucl.ac.uk/spm/; version DCM12.5 revision 7479). The log-family evidence was computed using the MACS toolbox (https://github.com/JoramSoch/MACS/releases/tag/v1.3), and Bayesian model comparisons were performed with the VBA toolbox (https://mbb-team.github.io/VBA-toolbox/). Codes for implementing model falsification, parameter and model recovery, as well as computational DCM to study predictive control, is available on GitHub (https://github.com/PierreGagnepain/predictive_control).

References

1. Stein, M. B. & Paulus, M. P. Imbalance of Approach and Avoidance: The Yin and Yang of Anxiety Disorders. *Biol. Psychiatry* **66**, 1072–1074 (2009).

2. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).

3. Homan, P. *et al.* Neural computations of threat in the aftermath of combat trauma. *Nat. Neurosci.* **22**, 470–476 (2019).

4. Brown, V. M. *et al.* Associability-modulated loss learning is increased in posttraumatic stress disorder. *eLife* **7**, e30150 (2018).

5. Gagne, C., Dayan, P. & Bishop, S. J. When planning to survive goes wrong: predicting the future and replaying the past in anxiety and PTSD. *Curr. Opin. Behav. Sci.* **24**, 89–95 (2018).

6. Seriès, P. Post-traumatic stress disorder as a disorder of prediction. *Nat. Neurosci.* **22**, 334–336 (2019).

7. Lissek, S. & van Meurs, B. Learning models of PTSD: Theoretical accounts and psychobiological evidence. *Int. J. Psychophysiol.* **98**, 594–605 (2015).

8. Dunsmoor, J. E. & Paz, R. Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biol. Psychiatry* **78**, 336–343 (2015).

9. Ehlers, A., Hackmann, A. & Michael, T. Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory* **12**, 403–415 (2004).

10. Mary, A. *et al.* Resilience after trauma: The role of memory suppression. *Science* **367**, (2020).

11. Gagnepain, P., Henson, R. N. & Anderson, M. C. Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proc. Natl. Acad. Sci.* **111**, E1310–E1319 (2014).

12. Brewin, C. R., Gregory, J. D., Lipton, M. & Burgess, N. Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. *Psychol. Rev.* **117**, 210–232 (2010).

13. Braver, T. S. The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.* **16**, 106–113 (2012).

14. Anderson, M. C., Bunce, J. G. & Barbas, H. Prefrontal-hippocampal pathways underlying inhibitory control over memory. *Neurobiol. Learn. Mem.* **134 Pt A**, 145–161 (2016).

15. Jiang, J., Heller, K. & Egner, T. Bayesian modeling of flexible cognitive control. *Neurosci. Biobehav. Rev.* **46**, 30–43 (2014).

16. Jiang, J., Wagner, A. D. & Egner, T. Integrated externally and internally generated task predictions jointly guide cognitive control in prefrontal cortex. *eLife* **7**, e39497 (2018).

17. Daunizeau, J. *et al.* Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. *PLOS ONE* **5**, e15554 (2010).

18. Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F. & Banich, M. T. The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive, Emotional, and Motor Processes. *Cereb. Cortex* **26**, 1634–1646 (2016).

19. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement. in *Classical Conditioning II: Current Research and Theory* 64–69 (New York: Appleton-Century-Crofts, 1972).

20. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **82**, 35–45 (1960).

21. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A Bayesian Foundation for Individual Learning Under Uncertainty. *Front. Hum. Neurosci.* **5**, (2011).

22. Palminteri, S., Wyart, V. & Koechlin, E. The Importance of Falsification in Computational Cognitive Modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).

23. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**, e49547 (2019).

24. Jacoby, L. L., Lindsay, D. S. & Hessels, S. Item-specific control of automatic processes: Stroop process dissociations. *Psychon. Bull. Rev.* **10**, 638–644 (2003).

25. Gagnepain, P., Hulbert, J. & Anderson, M. C. Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories. *J. Neurosci. Off. J. Soc. Neurosci.* **37**, 6423–6441 (2017).

26. Grisanzio, K. A. *et al.* Transdiagnostic Symptom Clusters and Associations With Brain, Behavior, and Daily Function in Mood, Anxiety, and Trauma Disorders. *JAMA Psychiatry* **75**, 201–209 (2018).

27. Konecky, B., Meyer, E. C., Kimbrel, N. A. & Morissette, S. B. The Structure of DSM-5 Posttraumatic Stress Disorder Symptoms in War Veterans. *Anxiety Stress Coping* **29**, 497–506 (2016).

28. Jammalamadaka, S. R. & Sengupta, A. *Topics In Circular Statistics-vol 5*. (World Scientific, 2001).

29. Desmedt, A., Marighetto, A. & Piazza, P.-V. Abnormal Fear Memory as a Model for Posttraumatic Stress Disorder. *Biol. Psychiatry* **78**, 290–297 (2015).

30. Kube, T., Berg, M., Kleim, B. & Herzog, P. Rethinking post-traumatic stress disorder – A predictive processing perspective. *Neurosci. Biobehav. Rev.* **113**, 448–460 (2020).

31. Henson, R. What can functional neuroimaging tell the experimental psychologist? *Q. J. Exp. Psychol. Sect. A* **58**, 193–233 (2005).

32. Perri, R. L. Is there a proactive and a reactive mechanism of inhibition? Towards an executive account of the attentional inhibitory control model. *Behav. Brain Res.* **377**, 112243 (2020).

33. Criaud, M., Wardak, C., Ben Hamed, S., Ballanger, B. & Boulinguez, P. Proactive Inhibitory Control of Response as the Default State of Executive Control. *Front. Psychol.* **3**, (2012).

34. Lyoo, I. K. The Neurobiological Role of the Dorsolateral Prefrontal Cortex in Recovery From Trauma: Longitudinal Brain Imaging Study Among Survivors of the South Korean Subway Disaster. *Arch. Gen. Psychiatry* **68**, 701 (2011).

35. Siehl, S., King, J. A., Burgess, N., Flor, H. & Nees, F. Structural white matter changes in adults and children with posttraumatic stress disorder: A systematic review and meta-analysis. *NeuroImage Clin.* **19**, 581–598 (2018).

36. van Belle, J., Vink, M., Durston, S. & Zandbelt, B. B. Common and unique neural networks for proactive and reactive response inhibition revealed by independent component analysis of functional MRI data. *NeuroImage* **103**, 65–74 (2014).

37. Czéh, B. *et al.* Chronic stress reduces the number of GABAergic interneurons in the adult rat hippocampus, dorsal-ventral and region-specific differences. *Hippocampus* **25**, 393–405 (2015).

38. Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescot, A. P. & Anderson, M. C. Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nat. Commun.* **8**, 1311 (2017).

39. Eshel, N. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).

40. Moutoussis, M., Shahar, N., Hauser, T. U. & Dolan, R. J. Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. *Comput. Psychiatry* **2**, 50–73 (2018).

41. Wenzlaff, R. M. & Wegner, D. M. Thought Suppression. *Annu. Rev. Psychol.* **51**, 59–91 (2000).

42. Hulbert, J. C., Henson, R. N. & Anderson, M. C. Inducing amnesia through systemic suppression. *Nat. Commun.* **7**, 11003 (2016).

43. Benoit, R. G. & Anderson, M. C. Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories. *Neuron* **76**, 450–460 (2012).

44. Sinclair, A. H. & Barense, M. D. Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends Neurosci.* **42**, 727–739 (2019).

45. Antony, J. W., Ferreira, C. S., Norman, K. A. & Wimber, M. Retrieval as a Fast Route to Memory Consolidation. *Trends Cogn. Sci.* **21**, 573–576 (2017).

46. Joo, H. R. & Frank, L. M. The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. *Nat. Rev. Neurosci.* **19**, 744–757 (2018).

47. Barron, H. C., Auksztulewicz, R. & Friston, K. Prediction and memory: A predictive coding account. *Prog. Neurobiol.* **192**, 101821 (2020).

48. Barron, H. C., Vogels, T. P., Behrens, T. E. & Ramaswami, M. Inhibitory engrams in perception and memory. *Proc. Natl. Acad. Sci.* 201701812 (2017) doi:10.1073/pnas.1701812114.

49. Brancu, M. *et al.* Subthreshold posttraumatic stress disorder: A meta-analytic review of DSM-IV prevalence and a proposed DSM-5 approach to measurement. *Psychol. Trauma Theory Res. Pract. Policy* **8**, 222–232 (2016).

50. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association, 2013). doi:10.1176/appi.books.9780890425596.

51. Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K. & Domino, J. L. The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and Initial Psychometric Evaluation: Posttraumatic Stress Disorder Checklist for *DSM-5*. *J. Trauma. Stress* **28**, 489–498 (2015).

52. de Berker, A. O. *et al.* Computations of uncertainty mediate acute stress responses in humans. *Nat. Commun.* **7**, 10996 (2016).

53. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat. Neurosci.* **3**, 1218–1223 (2000).

54. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex N. Y. N 1991* **26**, 3508–3526 (2016).

55. Friston, K. J., Mechelli, A., Turner, R. & Price, C. J. Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics. *NeuroImage* **12**, 466–477 (2000).

56. Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J. & Penny, W. Variational free energy and the Laplace approximation. *NeuroImage* **34**, 220–234 (2007).

57. Friston, K. J. Functional and Effective Connectivity: A Review. *Brain Connect.* **1**, 13–36 (2011).

58. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies — Revisited. *NeuroImage* **84**, 971–985 (2014).

59. Soch, J., Haynes, J.-D. & Allefeld, C. How to avoid mismodelling in GLM-based fMRI data analysis: cross-validated Bayesian model selection. *NeuroImage* **141**, 469–489 (2016).

60. Daunizeau, J., Adam, V. & Rigoux, L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLOS Comput. Biol.* **10**, e1003441 (2014).

61. Penny, W. D. *et al.* Comparing Families of Dynamic Causal Models. *PLoS Comput. Biol.* **6**, e1000709 (2010).

62. Wetzels, R. *et al.* Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 *t* Tests. *Perspect. Psychol. Sci.* **6**, 291–298 (2011).

63. Berens, P. CircStat: A Matlab Toolbox for Circular Statistics. *J. Stat. Softw.* **31**, 1–21 (2009).

Acknowledgements

We thank all the people who volunteered to take part in this study and the victim associations that supported this project. We thank the medical doctors (especially M. Mialon and E. Duprey) and the staff at the Cyceron biomedical imaging platform in Caen. We also thank the researchers; psychologists M. Deschamps, P. Billard, B. Marteau, R. Coppalle, and C. Becquet; technicians; and administrative staff at U1077 (Caen), at "Programme 13-Novembre" in Paris, at INSERM "Délégation Régionale Nord-Ouest" (Lille) and "Pôle Recherche Clinique"(Paris). We thank Jean-François Démonet for comments and feedbacks on this manuscript. We thank Elizabeth Portier for final English editing of the manuscript.

Funding: This study was funded by the French Commissariat-General for Investment (CGI) via the National Research Agency (ANR) and the "Programme d'investissement pour l'Avenir (PIA)." The study was realized within the framework of "Programme 13-Novembre" (EQUIPEX Matrice) headed by D.P. and F.E. This program is sponsored by the CNRS and INSERM and supported administratively by HESAM Université, bringing together 35 partners (see www.memoire13novembre.fr). G.L. is funded by a Ph.D. fellowship from the Normandy Region and Normandy University.

Author contributions

J.D., D.P., F.E., and P.G. designed the study. P.G. and G.L. conceptualized and implemented the computational model. P.G., J.D., D.P., F.E. obtained the financial support. C.P., A.M., and T.V. performed the data acquisition and F.F. managed and coordinated the research activity planning and execution. F.V. and V.d.L.S. supervised the MRI data collection and medical interviews. V.d.L.S. supervised the medical aspects of the study, and J.D. supervised the SCID interviews and psychiatric examinations. G.L. and P.G. analyzed the behavioral and functional data. G.L. and P.G. wrote the original draft. P.G. supervised the research. All the authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

8. Second study

Plasticity in control and memory circuits forecasts remission from PTSD

Giovanni Leone[1], Charlotte Postel[1], Florence Fraisse[1], Thomas Vallée[1], Fausto Viader[1], Vincent de La Sayette[1], Denis Peschanski[2], Jaques Dayan[1,3], Francis Eustache[1], Pierre Gagnepain[1]*

[1] Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, GIP Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France.

[2] Université Paris I Panthéon Sorbonne, HESAM Université, EHESS, CNRS, UMR8209, Paris, France.

[3] Pôle Hospitalo-Universitaire de Psychiatrie de l'Enfant et de l'Adolescent, Centre Hospitalier Guillaume Régnier, Université Rennes 1, 35700 Rennes, France.

***Corresponding author email**: pierre.gagnepain@inserm.fr

Pierre Gagnepain

GIP Cyceron, Boulevard Becquerel

14074, Caen, France

Tel.: +33 (0)2 314 701 59

Abstract

Individuals developing post-traumatic stress disorder (PTSD) following a traumatic experience present a memory disorder rooted in the alteration of hippocampal structures. Recently, it has been proposed that the persistence of intrusive memories in PTSD may additionally be linked in the disruption of the balance between predictive and reactive control of intrusive memories. We used a longitudinal design to investigate how these disruptions, central to PTSD, evolved in individuals exposed to the 2015 Paris terrorist attacks. We found that the remission from PTSD, three years after the trauma, was associated with increased left CA1 and right CA2-3/DG volumes and recovered memory control balance, compared to one year after. In stable PTSD, however, predictive control remains disrupted, while atrophy of the left CA2-3/DG progressed. Furthermore, increases reactive memory control and in hippocampal volumes forecasted the reduction of symptoms severity five years after the attacks, including intrusive re-experiencing. These findings reveal that neurocognitive plasticity occurring in control and memory circuits is central to understand the persistence of the trauma and remission process.

Introduction

Post-traumatic stress disorder (PTSD) has long been characterized as a disorder of memory rooted in the alteration of the hippocampus (van der Kolk, 2007; van Marle, 2015). More recently, PTSD has also been linked to a disorder of inhibitory functions that normally support the control of memory and reduce the accessibility of unwanted memory (Mary et al., 2020). However, little is known about the mechanisms underlying the eventual remission from PTSD and the evolution of these dysfunctions.

Vivid and distressing intrusive memories of the trauma, together with maladaptive avoidance of trauma reminder, are central features of PTSD. On the one hand, popular models link PTSD to a disorder of memory, rooted in the alteration of hippocampal functions and related structures (Brewin et al., 2010). According to this model, hippocampal dysfunction prevents the contextual integration of the traumatic engram. Contextual cues weakly connected to the trauma can trigger the involuntarily recall of the trauma (Bisby & Burgess, 2017). Poor contextual integration also gives the impression that the event is happening again in the present (Speckens et al., 2007), and may further lead to overgeneralization of fear (Steiger et al., 2015). Alteration of the hippocampus also impairs the extinction or updating of the original traumatic engram (Liberzon & Abelson, 2016), promoting persistence of the symptoms. This lack of contextual integration could be rooted in the impairment of pattern completion and separation mechanisms operated by the hippocampus (Kheirbek et al., 2012). These mechanisms normally allow the efficient indexation of cortical traces contacted by contextual cues entering the hippocampus (E. Rolls, 2013; E. T. Rolls, 2010), but would be impaired by the reduction of hippocampal volumes (Carr et al., 2010). In agreement with this view, we recently observed that the reduction of volume of Cornu Ammonis 1 (CA1), a subregion of the hippocampus performing pattern completion, was linked to the intrusive re-experiencing in individuals with PTSD, while the reduction of the CA2-3/Dentate Gyrus (DG) hippocampal subfield, central for pattern separation, was linked to avoidance behaviors (Postel et al., 2021).

On the other hand, we recently proposed that the persistence of the trauma may additionally be rooted in the generalized disruption of the memory control system (Mary et al., 2020). We implemented neutral and inoffensive intrusive memories in the laboratory in a group of

individuals exposed to the 2015 Paris terrorist attacks and nonexposed individuals. While reexperiencing these intrusive memories, analyses of brain connectivity revealed that nonexposed individuals and exposed individuals without PTSD could adaptively suppress memory activity, but exposed individuals with PTSD could not. In a follow-up study, we suggested that the alteration of brain connectivity during memory suppression reflected a new pathological mechanism of PTSD, rooted in the relationship between the brain's predictive and control mechanisms (Leone et al., under review). Using the Hierarchical Gaussian Filter (HGF, Mathys et al., 2011) to model memory intrusiveness during the suppression task according to a meta-Bayesian framework (Daunizeau et al., 2010), we tracked the trial-by-trial hidden beliefs about the probability of experiencing intrusive memories. We then hypothesized that, similarly to other forms of control (Anderson et al., 2016; Braver, 2012), memory suppression imply the coordinated equilibration of two control mechanisms. A predictive control based on the expected probability of experiencing intrusions, and a reactive correction when intrusive memories enter consciousness. We used dynamic causal modeling (DCM, Friston et al., 2003) to investigate predictive and reactive control of intrusive memories. To understand the influence of intrusive beliefs and related prediction errors (PE), we incorporated these computational indexes as modulators of the down-regulation of the hippocampus and the precuneus orchestrated by the middle frontal gyrus (MFG). We found that, contrarily to resilient and nonexposed individuals, participants with PTSD showed imbalance between predictive and reactive control over the hippocampus. This imbalance was characterized by exaggerated predictive control and a lack of reactive control when intrusive memories needed to be purged, and was correlated with the severity of intrusive and avoidance symptoms.

Despite studies attempted to identify dysfunctions associated with the risk of developing PTSD following a traumatic experience, the origin of hippocampal alterations and inhibitory dysfunctions remains largely unknown and is still debated. On the one hand, hippocampal and memory control disorders could precede the trauma, constituting a pre-existing risk factor. Supporting of this view, one study involving identical twins discordant for trauma exposure have suggested that hippocampal atrophy pre-exists before the trauma (Gilbertson et al., 2002). Furthermore, several studies have identified multiple genetic, behavioral and neural factors partially explaining the individual variability of executive functions (Friedman & Miyake, 2017), including the ability to suppress unwanted memories (Levy & Anderson, 2008) in the healthy population.

On the other hand, hippocampal atrophy and memory control deficits could reflect stress-induced alterations following the trauma. Animal models have shown that stress can induce neurotoxic effects on the hippocampus via the production of glucocorticoids and excessive glutamate, resulting in neuronal and synaptic loss (Gao et al., 2014; Gould, 2007; McEwen et al., 2016a; Schoenfeld et al., 2017). A prospective study measuring hippocampal volume in healthy soldiers before and after the military service have shown that the development of PTSD-related symptoms was correlated with reductions in hippocampal volumes, and not to the initial volumes (Admon et al., 2013). Furthermore, people remitted from PTSD show greater hippocampal volumes compared to veterans with ongoing PTSD, possibly suggesting that the remission from PTSD is accompanied by some forms of plasticity of the hippocampus (Apfel et al., 2011). The cascade of molecular alterations related to stress also impairs executive functions, resulting in the reduction of the prefrontal cortex functioning and the consequential deficits in inhibitory control, working memory and attention (Arnsten, 2009). However, little is known about the effect of stress on memory suppression.

In the current study, we used a longitudinal experimental design to investigate the relationship between the evolution in time of both memory control dysfunctions and hippocampal subfields' volumes on the one hand, and the clinical evolution in the individuals exposed to the Paris November 2015 terrorist attacks on the other hand. Hippocampal subfields volumes, fMRI activity during the TNT task, and PTSD clinical interview for diagnosis, were recorded 6 to 18 months at time 1 (T1), and 30 to 42 months at time 2 (T2), after the traumatic event (2016 and 2018, respectively). A further follow-up phone interview provided the measurement of symptoms' severity 60 to 62 months after the trauma (T3, 2020, see **Figure 1a**). At both T1 and T2, we processed hippocampal subfields as in Postel et al., (2021) and we used computational modeling and DCM to quantify the brain mechanisms of predictive and reactive control of intrusive memories, as in Leone et al. (under review). The exposed group was composed of individuals suffering from partial or full PTSD symptoms both at T1 and T2 (denoted Stable PTSD), individuals recovering from a PTSD at T1 (denoted Remitted PTSD), and individuals showing no noticeable impairment after the trauma both at T1 and T2 (denoted Stable non-PTSD). We tested the relationship between the evolution of predictive and reactive control and the evolution of CA1 and CA2-CA3-DG volumes and the future evolution of symptoms' severity, focusing on intrusions and avoidance. According to a pre-existing account of control or memory disorder of PTSD, inhibitory dysfunctions or hippocampal alterations that we observed at T1 should still be present at T2, even in

individuals remitted from PTSD. According to a stress-induced account, however, some of the dysfunctions or alteration observed at T1 may not persist and recover at T2, along with reduction of critical symptoms.

Results

Participants

After exclusions and attritions, the final sample was composed by 110 trauma-exposed and 75 nonexposed participants. Crucially, within the exposed participants, 59 met the criteria for complete or partial PTSD at T1 and 51 did not. DCM data were available from 173 participants at T1 and 162 participants at T2 and Hippocampal Subfield (HS) acquisition and segmentation were effective in 147 participants at T1 and 146 participants T2 (see **Table 1** for detailed information about the data availability and demographics within the included sample). Participants were also invited to fill online questionnaires and structured interviews within May and June 2020 (T3). At all the three phases, symptoms' severity was assessed with the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5, Blevins et al., 2015).

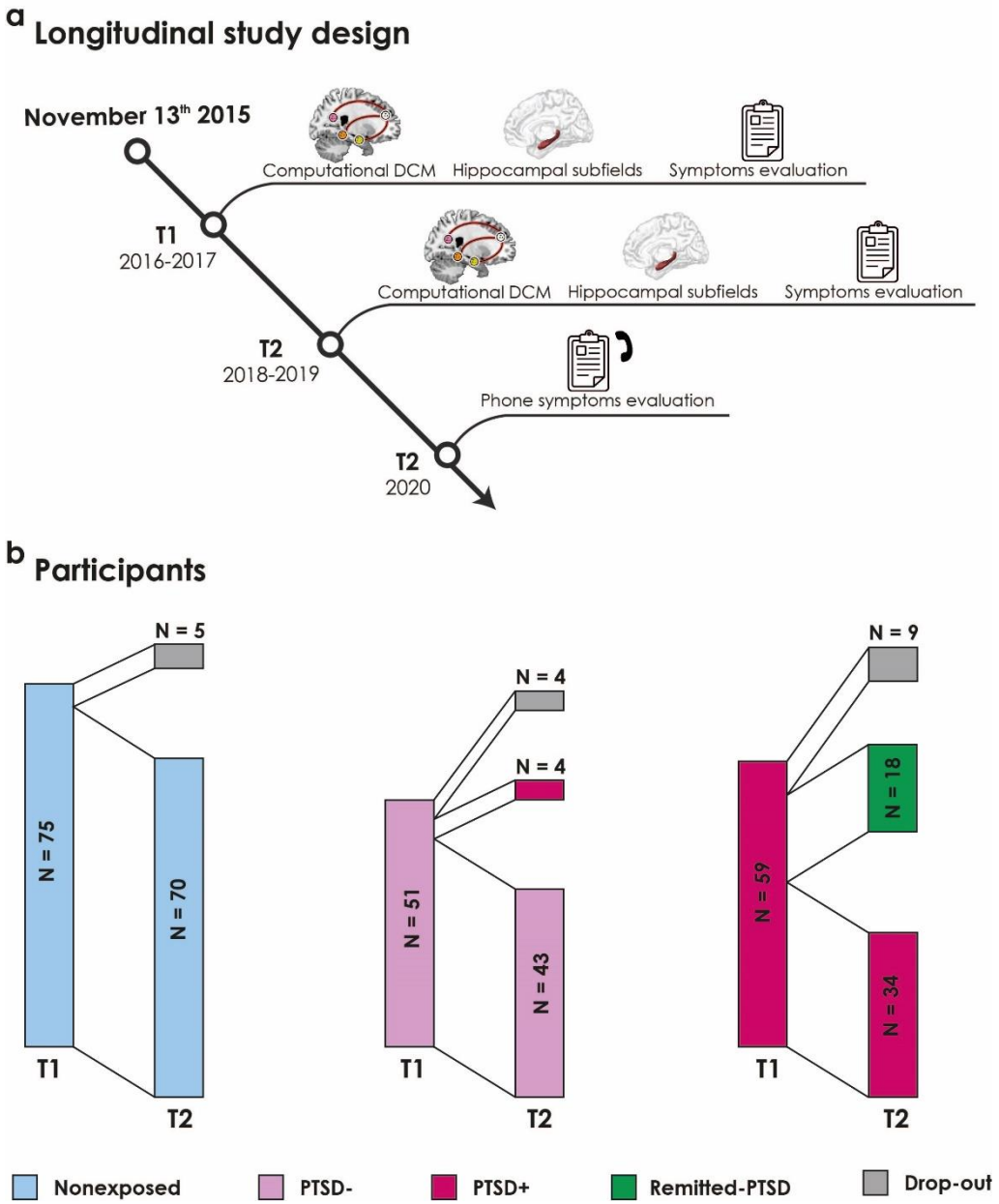
Group	N (male)	Age	DCM T1 (N)	DCM T2 (N)	DCM T1 & T2 (N)	Subfields T1 (N)	Subfields T2 (N)	Subfields T1 & T2 (N)
Non exposed	75 (34)	33.83 ± 11.43	72	67	65	56	57	45
PTSD-	51 (31)	36.83 ± 6.95	46	46	42	39	41	34
PTSD+	59 (27)	37.26 ± 8..23	55	49	45	52	48	44
Total	186 (92)	35.76 ± 9.46	173	162	152	147	146	123

Table 1. Data availability. Demographics and number of participants for each modality and time, as well as within each modality across time.

Clinical evolution over time

We first investigated how the diagnosis of PTSD evolved after two years from the first data acquisition. Exposed participants were diagnosed using the structured clinical interview for DSM-5 (SCID) conducted by an expert psychologist (Zlotnick et al., 2002). We included in

the PTSD group participants meeting the criteria for the diagnosis of full or partial PTSD (for a detailed description of the inclusion procedure, see Mary et al., 2020). Within the 52 exposed participants without PTSD (PTSD-) at T1, 36 participants were stable, seven showed only intrusive symptoms without any functional impairment, four participants reached criteria for PTSD and four participants dropped from the study. The seven participants presenting



only intrusive symptoms and that were not diagnosed as full or partial PTSD, were included in the "stable non-PTSD" group. Concerning the PTSD+ group, 34 out of 59 participants remained stable at T2, 18 participants remitted from PTSD ("remitted-PTSD" group) and nine dropped out from the study (See **Figure 1b**). Seventy out of 75 nonexposed participants joined the second phase of the study.

Fig. 1. a) Longitudinal study design. b) Group sizes and clinical changes in time.

The analyses focused on the following four groups: nonexposed, stable non-PTSD, remitted PTSD and stable PTSD. Despite representing interesting cases of late-onset PTSD, the four participants who developed PTSD at T2 were excluded from the current study because of insufficient group size to perform statistical analyses. Further case-studies would be needed to better understand the risk factors associated with developing PTSD later in time in initially resilient individuals.

Longitudinal changes in predictive and reactive control of intrusive memories

We first questioned how predictive and reactive control evolved in time in our groups. For each group we tested Control*Time interactions by comparing through t-tests the balance between predictive and reactive control at T2 to the balance at T1. For each time point, the balance was calculated by subtracting reactive control (i.e., PE modulation) from predictive control (i.e., beliefs modulation). We focused on the coupling parameters connecting MFG to the whole hippocampus (wHIP) and the precuneus (PC).

We found a Time*Control interaction in the wHIP in the remitted-PTSD group ($T = 2.415$, $p = 0.029$, $df = 15$, 95% CI = [0.08 1.23]). This interaction was characterized by the reduction of the imbalance in memory control at T2 (i.e. no significant difference between predictive and reactive control; $T = -1.206$, $p = 0.245$, $df = 16$, 95% CI = [-1.13 .31]), compared with T1, which was, on the contrary, associated with a significant imbalance in favor of predictive control ($T = -4.099$, $p < 0.001$, 95% CI = [-1.01 -0.32]). No Time*Control effect in the downregulation of the wHIP was observe in the other three, stable, groups (see **Table 2** and **Figure 2**). Concerning the PC, we found a Time*Control interaction in the stable non-PTSD group ($T = 2.404$, $p = 0.021$, $df = 38$ 95% CI = [0.10 1.18]). This interaction was characterized by a greater balance between predictive and reactive control at T2 ($T = -1.423$, $p = 0.162$, $df = 41$, 95% CI = [-0.63 0.11]) compared T1 was imbalanced towards predictive control ($T = -3.406$, $p = 0.001$, $df = 38$, 95% CI = [-1.29 -.033]). This interaction was not observed in the other three groups.

Altogether, these results showed that the imbalance between predictive and reactive control that characterized T1 disappeared at T2 specifically in remitted-PTSD. On the contrary, the stable resilient group, while maintaining the memory control balance over the hippocampus, showed a reduction of the imbalance over the precuneus between T1 and T2.

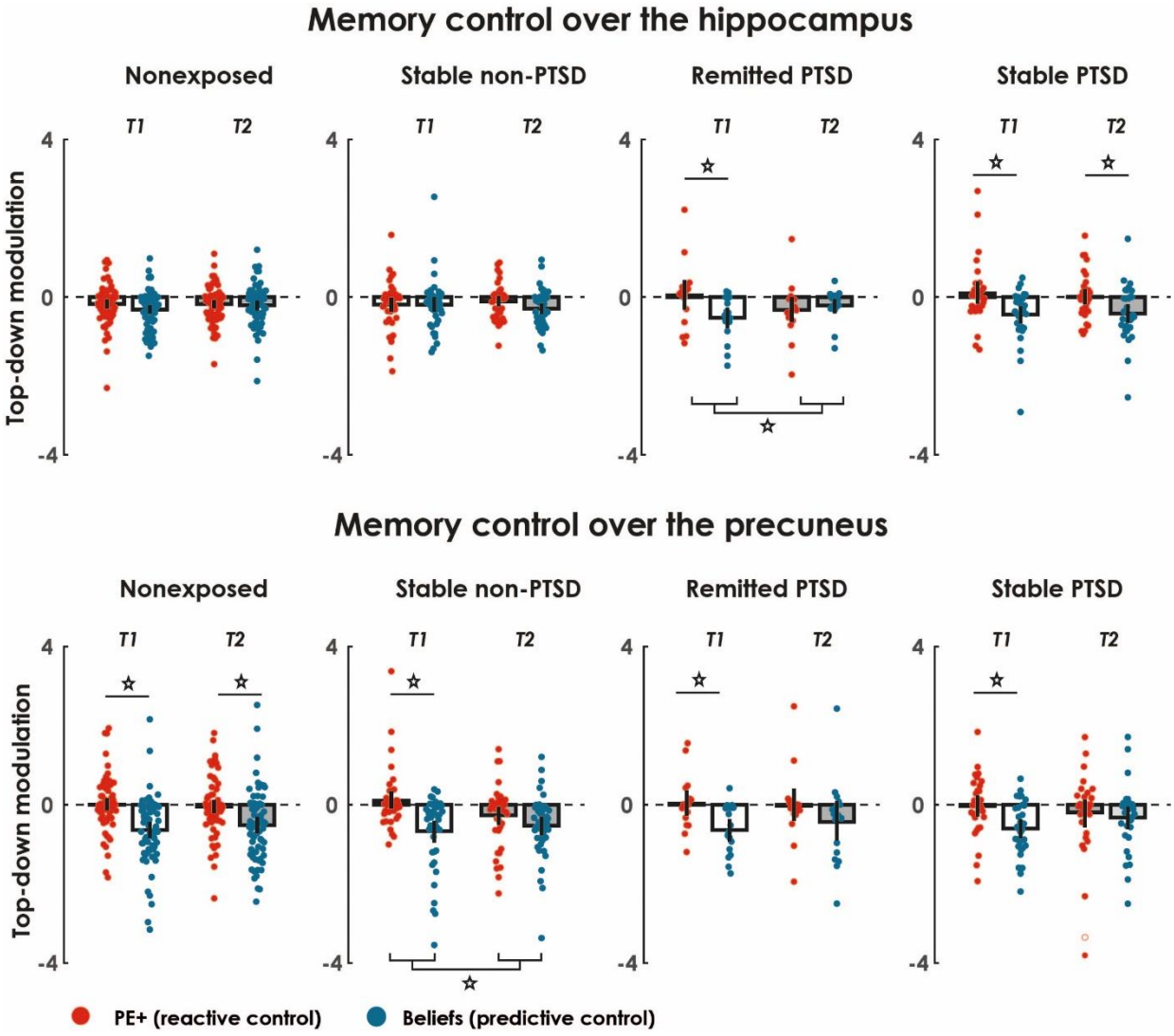


Figure 2. . Top-down coupling parameters during belief- and PE-driven suppression, respectively representing predictive and reactive control, respectively, at T1 and T2. Red and blue circles represent the modulation of the top-down coupling between the MFG and PE+ and belief target regions (hippocampus and precuneus). Error bars represent the bootstrapped 95% CI of the group mean.

Table 2 Longitudinal changes in memory control balance					
Group	ROI	T	DF	p	95% CI
Nonexposed	wHIP	0.740	64	0.462	[-0.17 0.37]
	PC	0.854	64	0.396	[-0.22 0.56]
Stable non-PTSD	wHIP	-1.155	38	0.255	[-0.59 0.16]
	PC	2.404	38	0.021*	[0.10 1.18]
Remitted PTSD	wHIP	2.415	15	0.029*	[0.08 1.23]
	PC	-0.037	15	0.971	[-0.67 0.65]
Stable PTSD	wHIP	0.525	28	0.604	[-0.45 0.77]
	PC	1.204	28	0.239	[-0.27 1.05]

Table 2. Longitudinal changes in memory control balance. Within-group t-tests comparing the imbalance at T2 to the imbalance at T1. ROI, region of interest; DF, degrees of freedom; 95% CI, 95% confidence intervals.

Longitudinal changes in hippocampal subfields volumes

We investigated whether there were longitudinal changes in HS volumes for each specific group. We computed the percentage of change of adjusted volumes between T1 and T2 (i.e. T2-T1/T1) for both CA1 ($\Delta\%_{CA1(T2-T1)}$) and CA2-3/DG ($\Delta\%_{CA2\text{-}3/DG(T2-T1)}$). For each group, one-sample two-tailed t-tests were used to test the hypothesis that $\Delta\%_{CA1(T2-T1)}$ and $\Delta\%_{CA2\text{-}3/DG(T2-T1)}$ were different from zero, that is, to investigate whether there were significant changes in HS volumes between T1 and T2.

In the left hemisphere, we found a significant increase in CA2-3/DG in the remitted PTSD group (T = 2.99, p = 0.009, df = 15, 95% CI = [0.01 .04]), but no other significant effect in the other groups (see **Table 3** and **Figure 3**). In the right hemisphere, we observed a significant increase in the volume of CA1 in the remitted PTSD group (T = 2.31, p = 0.035, df = 15, 95% CI = [0.003 .03]). In the stable PTSD group, however, a significant reduction in the volume of the right CA2-3/DG was found (T = -3.33, p = 0.0025, df = 27, 95% CI = [-0.033 -0.009]). Further statistical details are reported in **Table 3**.

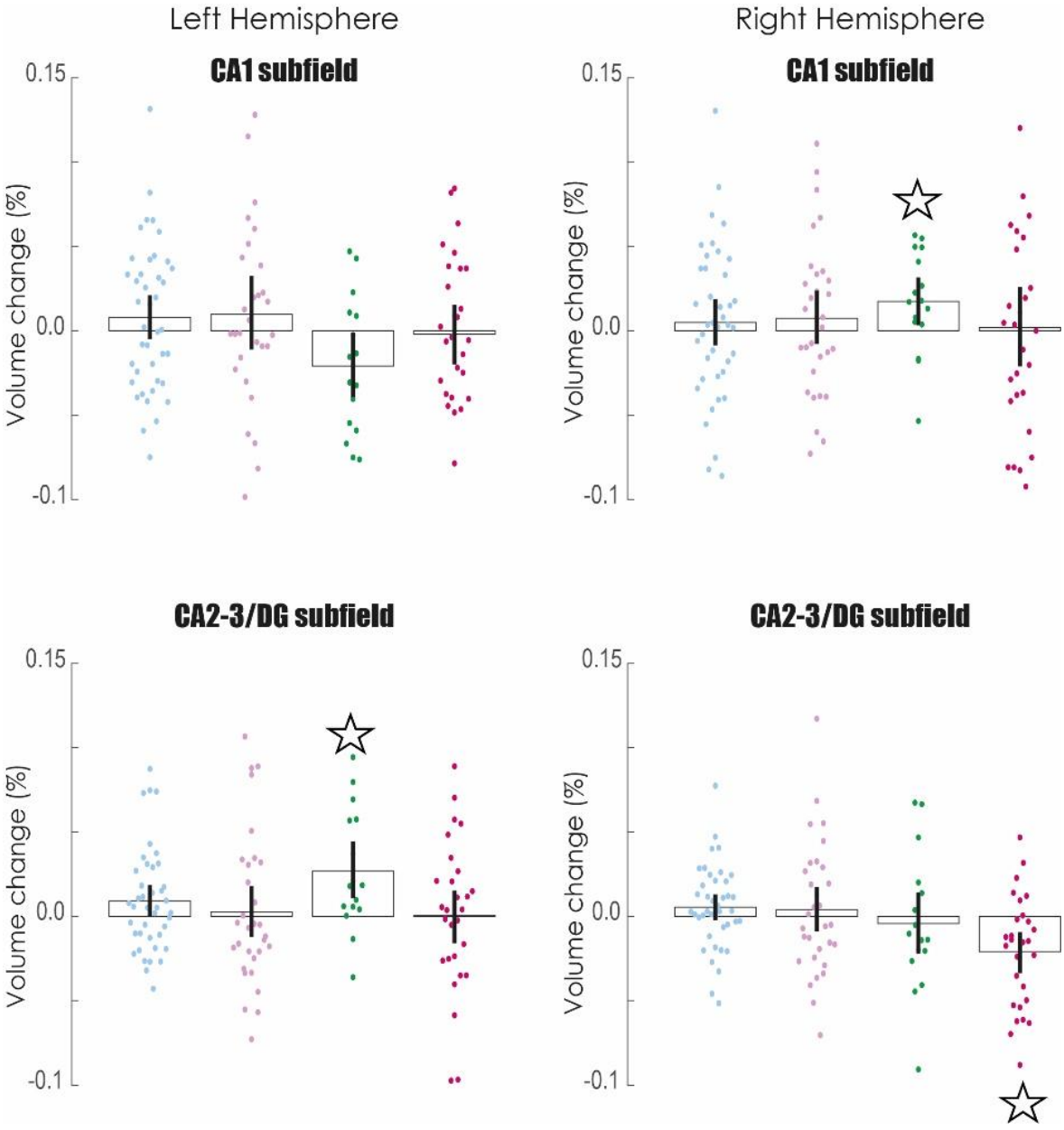


Figure 3. Percentage of longitudinal changes of the hippocampal volumes. Error bars reflect bootstrapped 95% CI and hence indicate significant atrophy when both extremes are below zero.

	Table 3 Longitudinal changes in hippocampal volumes					
Group		**ROI**	**T**	**DF**	**p**	**95% CI**
Nonexposed	Left	CA1	1,167	40,000	0,250	[-0,005 0,021]
		CA2-3/DG	1,902	40,000	0,064	[0 0,019]
	Right	CA1	0,713	40,000	0,480	[-0,009 0,019]
		CA2-3/DG	1,408	40,000	0,167	[-0,002 0,013]
Stable non-PTSD	Left	CA1	0,856	30,000	0,399	[-0,013 0,033]
		CA2-3/DG	0,341	30,000	0,735	[-0,013 0,019]
	Right	CA1	0,903	30,000	0,374	[-0,008 0,022]
		CA2-3/DG	0,540	30,000	0,593	[-0,01 0,018]
Remitted PTSD	Left	CA1	-2,076	15,000	0,055	[-0,04 -0,002]
		CA2-3/DG	2,993	15,000	0,009*	[0,011 0,044]
	Right	CA1	2,319	15,000	0,035*	[0,002 0,031]
		CA2-3/DG	-0,413	15,000	0,685	[-0,024 0,015]
Stable PTSD	Left	CA1	-0,235	27,000	0,816	[-0,019 0,015]
		CA2-3/DG	0,046	27,000	0,964	[-0,017 0,016]
	Right	CA1	0,171	27,000	0,865	[-0,021 0,026]
		CA2-3/DG	-3,332	27,000	0,003*	[-0,033 -0,009]

Table 3. Longitudinal changes in HS volumes. One-sample t-tests assessing whether percentage changes in adjusted HS volumes are different to zero. ROI, region of interest; DF, degrees of freedom; 95% CI, 95% confidence intervals.

Plasticity of control processes and hippocampal circuits forecast remission of the trauma.

We then investigated whether individual variations in control and hippocampal markers in the stable PTSD group, forecasted the changes in symptoms at T3. More specifically, we asked whether the improvement or degradation between T1 and T2 of hippocampal control mechanisms and volumes, were related to the evolution between T2 and T3 of the main

criteria of PTSD (i.e. intrusive re-experiencing, avoidance, negative alteration of cognition and mood, and arousal). Those changes in psychopathology were quantified using the PTSD checklist for DSM-5 (Blevins et al., 2015). This crucial analysis would allow indirectly inferring whether the reduction of hippocampal and memory control disorders is more likely to precede or to provoke PTSD remission.

Concerning control processes, we found that longitudinal changes in reactive control over the hippocampus forecasted changes in intrusive re-experiencing (Spearman's $rho_{skipped}$ = 0.40, p = 0.046, 95% bootstrapped CI [0.02 0.69]; **Figure 4**). To further characterize this significant relationship, we categorized individuals with stable PTSD according to the evolution of the reactive control over time. We created an "improvement" group, showing improved reactive inhibition at T2 compared with T1, and a "degradation" group, showing an alteration of this inhibitory function at T2 compared with T1. Planned comparisons revealed that the "improvement" group expressed fewer intrusions at T3 (t(9) = -2.95, p=.008; 95% bootstrapped CI = [-6.9 -1.65]), while this effect was absent in the "degradation" group (t(10) = -.46, p=.32; 95% bootstrapped CI = [-1.7 1.02]). No further significant correlations were observed with other PTSD dimensions or predictive control (see **Figure 4**).

Concerning hippocampal CA2-3/DG (averaged across left and right hemisphere for this analysis), we found that longitudinal changes in the volumes of this region forecasted changes in intrusive re-experiencing (Spearman's $rho_{skipped}$ = -.72, p = 0.002, 95% bootstrapped CI [-.92 -0.42]), and negative alteration of cognition and mood (Spearman's $rho_{skipped}$ = -.45, p = 0.019, 95% bootstrapped CI [-.79 -0.10]; **Figure 5**). To further characterize these significant relationships, we categorized individuals with stable PTSD according to the changes in hippocampal volume over time. We created a "plasticity" group, showing greater CA2-3/DG volume at T2 compared with T1, and an "atrophy" group, showing a reduction of this volume at T2 compared with T1. Planned comparisons revealed that the "plasticity" group expressed fewer intrusions at T3 (t(6) = -3.66, p=.005; 95% bootstrapped CI = [-7.1 -2.4]), while this effect was absent in the "atrophy" group (t(12) = 0, p=.5; 95% bootstrapped CI = [-1.03 1.2]). The "plasticity" group did not express further improvement in cognition and mood (t(9) = -1.26, p=.12; 95% bootstrapped CI = [-5.8 1.2]; see **Figure 5**), despite the presence of a significant relationship when all individuals are considered together.

Concerning hippocampal CA1 (averaged across left and right hemisphere for this analysis), we found that longitudinal changes in the volumes of this region forecasted changes in

intrusive re-experiencing (Spearman's $rho_{skipped}$ = -.57, p = 0.008, 95% bootstrapped CI [-.76 -0.26]), avoidance (Spearman's $rho_{skipped}$ = -.51, p = 0.002, 95% bootstrapped CI [-.73 -0.18]), and negative alteration of cognition and mood (Spearman's $rho_{skipped}$ = -.45, p = 0.01, 95% bootstrapped CI [-.67 -0.13]; **Figure 5**). These three significant dimensions were associated with significant improvement at T3 when "plasticity" individuals were considered separately from individuals associated with a longitudinal trajectory characterized by "atrophy" (Intrusion: t(11) = -2.05, p=.03; 95% bootstrapped CI = [-5.4 -.33]; Avoidance: t(11) = -3.09, p=.005; 95% bootstrapped CI = [-3.8 -1]; Cognition & Mood: t(11) = -2.63, p=.012; 95% bootstrapped CI = [-5.5 -1.1]; see **Figure 5**).

In summary, the plasticity between T1 and T2 of reactive control mechanisms is specifically related to a latter reduction in intrusive re-experiencing at T3. Hippocampal plasticity, however, seems involved in a broader spectrum of symptoms remission.

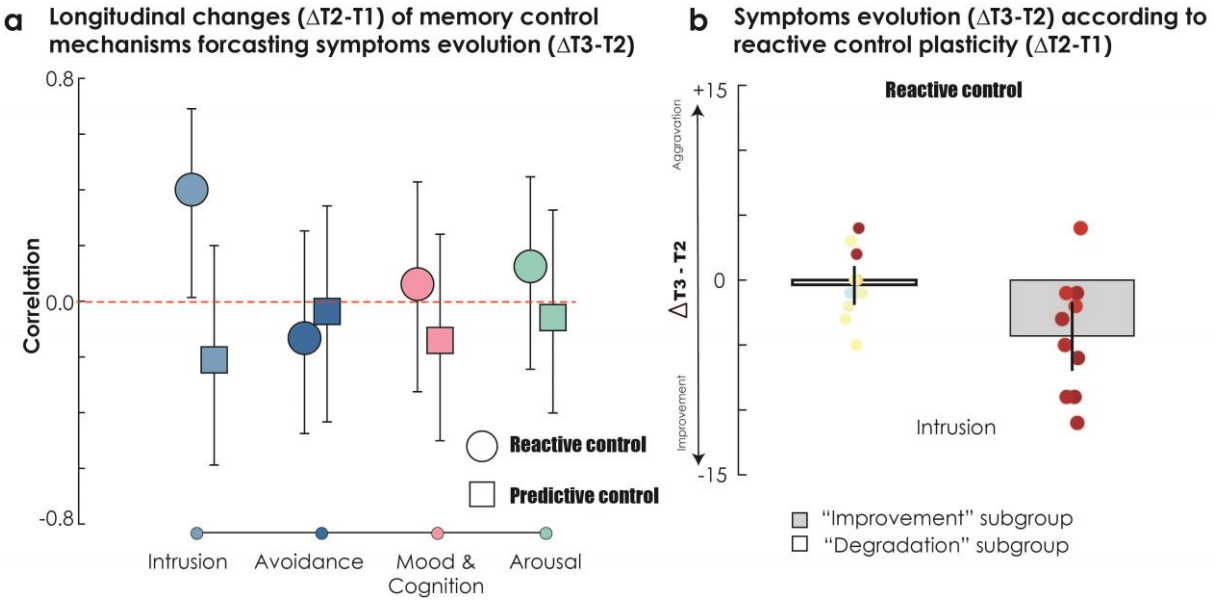


Figure 4. Relationships between longitudinal changes (ΔT2 – T1) of memory control processes and symptoms evolution at T3 (ΔT3 – T2). (**A**) Correlation in stable PTSD. Error bars reflect bootstrapped 95% CI of the correlation, and thus indicate significance when they do not overlap with 0. (**B**) Symptoms evolution at T3 according to individual variations in the plasticity of reactive control mechanisms. Error bars reflect bootstrapped 95% CI of the mean, and thus indicate significance when they do not overlap with 0.

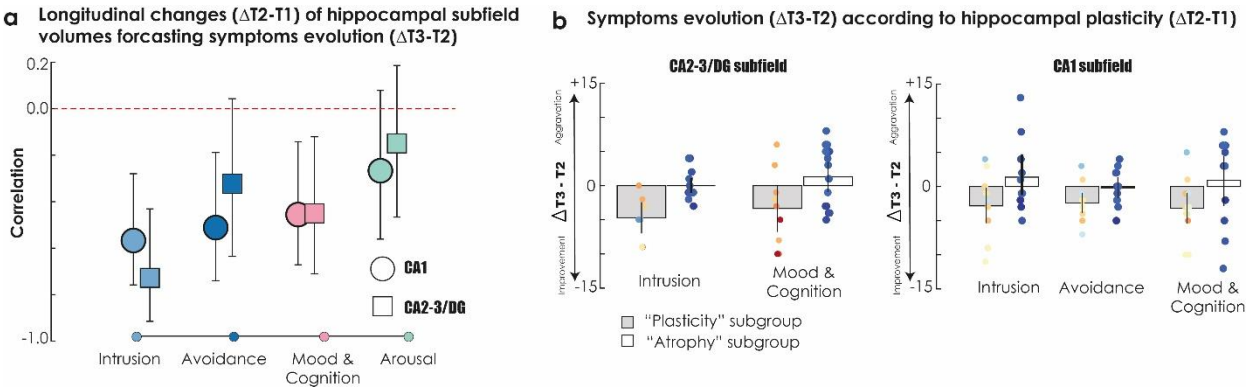


Figure 5. Relationships between longitudinal changes (ΔT2 – T1) of hippocampal volumes and symptoms evolution at T3 (ΔT3 – T2). (**A**) Correlation in stable PTSD. Error bars reflect bootstrapped 95% CI of the correlation, and thus indicate significance when they do not overlap with 0. (**B**) Symptoms evolution at T3 according to individual variations in hippocampal plasticity. Error bars reflect bootstrapped 95% CI of the mean, and thus indicate significance when they do not overlap with 0.

Discussion

PTSD has been described as a disorder of memory (Brewin et al., 2010; van Marle, 2015), characterized in Humans by hippocampal alteration and volume reduction (van der Kolk, 2007; van Marle, 2015). More recently, PTSD has additionally been linked to a disorder of control mechanisms (Leone et al., under review; Mary et al., 2020), normally altering the accessibility of unwanted memory traces. In line with these two views of PTSD, we reported, in two previous studies, that individuals with PTSD showed an imbalance between predictive and reactive control of intrusive memories (Leone et al., under review) together with an alteration of hippocampal CA1 and CA2-3/DG subfields (Postel et al., 2021). However, whether these hallmarks of PTSD reflect maladaptive outcomes following a traumatic experience immutable to changes is still unknown. Here, using a longitudinal experimental design, we found that individuals remitted from PTSD four years after a traumatic experience overcame this imbalance, mostly due to the gain of reactive control. These findings support recent proposal on the protective role of memory inhibition to alter the expression the traumatic memory traces (Leone et al., under review; Mary et al., 2020). In addition, individuals remitted from PTSD also showed hippocampal plasticity, while those with chronic and stable PTSD showed atrophy in the CA2-3/DG hippocampal subfield. Finally, the presence of neurocognitive plasticity between T1 and T2 measurements, characterized by

increased reactive inhibition and hippocampal volumes, were predictive of PTSD symptoms reductions at three years after the trauma.

Concerning control mechanisms, although these results alone cannot discern the temporal and causal relationship between memory control dysfunctions and PTSD, they support the hypothesis that these alterations may follow the trauma, rather than precede it. The association between the recovery in time of reactive memory control and remission from PTSD suggests that dysfunctions in control mechanisms might be an effect of intense stress and normalize when stress return to baseline levels. Stress can impair executive functions (Arnsten, 2009) by inducing the relocation of the executive resources normally supporting inhibition, working memory and flexibility towards the handling of the stressor (Shields et al., 2016). The release of glucocorticoids associated with intense stress cause changes in the glutamate neurotransmission in the PFC (Popoli et al., 2012), which have detrimental effects on the PFC-dependent cognitive functions (Qin et al., 2009; Yuen et al., 2012), including inhibitory control. Reductions in stress could explain the recovered balance in memory control mechanisms in remitted PTSD. Compatibly with this hypothesis, evidence has shown that blocking glucocorticoid receptors in the PFC improved executive functions (Butts et al., 2011) and stress reduction interventions improve executive functions (Moynihan et al., 2013).

Alternatively, the recovered ability to purge away intrusive memories in individuals remitted from PTSD may be mediated by the hippocampal GABAergic system. Lower hippocampal GABA concentrations are related to decreased PFC inhibition over the hippocampus and reduced forgetting (Schmitz et al., 2017). Animal studies have reported that prolonged stress induces a reduction in GABA hippocampal concentrations (Harvey et al., 2004) and alterations in $GABA_A$ receptors (Gunn et al., 2011). Alterations in the hippocampal GABAergic system could specifically impair reactive control. A recent model of memory suppression has proposed that predictive and reactive control could involve different neurobiological mechanisms (Anderson et al., 2016). According to this model, predictive control would target the entorhinal inputs to the hippocampus to gate its activity and reactive control would directly target the inhibition of the hippocampus. Thus, reactive control efficacy in targeting hippocampal activity can directly depend on the functioning of hippocampal GABA. Reductions in stress could restore the functioning of this hippocampal inhibitory system, reestablishing the hippocampal excitation/inhibition equilibrium, possibly contributing to the remission from PTSD.

Improved reactive memory control could be either a consequence or a cause of PTSD symptoms' severity fading, especially if such decrease is associated with a reduced stress. Crucially, we found that increased reactive control predicted future reductions in intrusive symptoms, suggesting a causal involvement of the ability to inhibit unwanted memory traces through the downregulation of hippocampal activity in the reduction of the accessibility of the traumatic engrams. Confirming these results, participants improving reactive control at T2 had significantly lower intrusive symptoms at T3 compared to participants showing a degradation of reactive control at T2. Retrieval entails consolidated memories entering in a vulnerable state (Kida, 2019; Schwabe et al., 2014), and the reactivation of a memory trace might be necessary condition to memory destabilization and active forgetting. It has been hypothesized that the reconsolidation of a memory trace depend on the intensity its reactivation, in a U-shaped relationship (Sinclair & Barense, 2019). Accordingly, moderated reactivations would weaken the memory traces. The intrusive memories successfully controlled via reactive control are transitory moderate reactivations of the unwanted memory engram, potentially facilitating the trace destabilization and weakening. Alternatively, it has been hypothesized that "inhibitory engrams" parallel the neuronal connections forming memories, silencing the activation of these excitatory engrams (Barron et al., 2016). In this context, reactive control of intrusive memories could potentiate the connections of these inhibitory engrams silencing specific unwanted memories.

Reductions of the volumes of HS have been reported in individuals developing PTSD following a traumatic experience but not in resilient individuals (Chen et al. 2018; Hayes et al. 2011; Postel et al. 2021). However, little is known about the causal link between HS and trauma and the relationship between evolution of HS volumes and remission from PTSD. We did find significant increase of right CA1 and left CA2-3/DG volumes in remitted PTSD. Furthermore, individuals with stable PTSD showed, overall, an atrophy of the right CA2-3/DG. However, the presence of hippocampal plasticity in this sample forecasted symptoms reductions at T3, including intrusion, alteration of mood and cognition, and avoidance. Thus, these findings suggest that the reduction of hippocampal volumes, in addition of being a pre-existing condition (Koch et al., 2021), is also exacerbated by stress-induced maladaptive response. However, individuals who manage to cope with this maladaptive atrophy, increase their chances of recovery from PTSD.

Chronic stress can impact hippocampal functioning and integrity by causing a prolonged exposure to glucocorticoids (Gagnon & Wagner, 2016). High stress-induced hippocampal

cortisol levels have been found to impair the hippocampal-dependent identification of threatening context in mice, inducing a pathological fear responses mimicking the behavior of Human PTSD (Kaouane et al., 2012). CA3 and the DG have been reported to be preferentially involved in pattern separation, a functional property allowing the hippocampus to separate overlapping memory traces (E. Rolls, 2013; Yassa & Stark, 2011). CA3 granule cells determine the separation of the two traces, which will be stored separately (Yassa & Stark, 2011). The DG is the only part of the hippocampus where neurogenesis is still possible at the adult age (Bergmann et al., 2015), playing a fundamental role in the contextual discrimination of overlapping memories (Surget & Belzung, 2021). CA1 is mostly implicated in pattern completion, a functional property allowing the hippocampus to retrieve memories basing on incomplete cues (E. Rolls, 2013). Animal studies have suggested that stress-related atrophy of this HS may be largely due to the loss of GABAergic interneurons (Czéh et al., 2015; McEwen et al., 2016b). Altogether, these studies suggest that reductions in the volumes of CA1 and CA2-3/DG subfield could cause the loss of contextual integration, leading to generalization of fear to non-threatening stimuli (Besnard & Sahay, 2016), avoidance and the persistence of intrusive memories. The reduction of stress levels accompanied by CA1 and CA2-3/DG plastic changes could precede and be causally involved in the future reduction of PTSD symptoms.

Partially contradicting our findings, a recent study did not found any correlation between longitudinal changes in HS volumes and PTSD symptoms' evolution (Weis et al., 2021). However, some substantial methodological differences may explain the discrepancy between the findings of the two studies. First, Weis et al. (2021) tested correlations between HS evolution and concomitant symptoms evolution, while we were interested in predictive analyses. Second, the authors collected HS data two weeks and six months after the trauma, a time window potentially too small to capture significant structural changes in the hippocampus. Third, the spatial resolution obtained with our specific MRI sequence is sensibly higher than with the standard MRI-T1 images used in Weis et al. (2021), and allow a more precise and reliable segmentation of the hippocampal subfield.

Altogether, our longitudinal design allowed us to show a causal involvement of memory control and hippocampal subfields' plasticity in determining future symptoms' evolution in PTSD. The recovery of reactive control was selectively predictive of future reductions in intrusive symptoms, and plastic changes in CA1 and CA2-3/DG were predictive of a more general reduction in the severity of PTSD. These results shed light on a dual mechanism

involving both memory control and hippocampal functioning restauration to inaugurate future clinical improvements, potentially representing a target for therapeutic interventions.

Material and methods

Participants

Two hundred participants were recruited in the framework of the longitudinal multidisciplinary project "13-Novembre" (https://www.memoire13novembre.fr), including 120 participants exposed to the Paris terrorist attacks and 80 nonexposed. Structural and functional MRI data were acquired in 2016 (T1) and in 2018 (T2), one year and three years after the terrorist attacks, respectively. At both time, PTSD symptoms' severity was also quantified with the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5, Blevins et al., 2015) and depression symptoms were examined using the Beck Depression Inventory (BDI, Beck et al., 1996). The study was approved by the regional research ethics committee ("Comité de Protection des Personnes Nord-Ouest III", sponsor ID: C16-13, RCB ID: 2016-A00661-50, clinicaltrial.gov registration number: NCT02810197). All participants gave written informed consent before participation, in agreement with French ethical guidelines. Detailed description of participants, inclusion criteria, materials and task procedure can be found in Leone et al. (under review), Mary et al., (2020) and Postel et al., (2019).

Think/No-Think, computational modeling and dynamic causal modeling

At both T1 and T2, participants performed the Think/No-Think task. Before the fMRI acquisition, participants intensively learned neutral French word object. We recorded fMRI activity during the TNT phase. During this task, 36 cue words repeated 8 times were displayed either in green (think condition) or red (no-think condition). During think trials, participants had to visualize and recall the associated object with as many details as possible. During no-think trials, participants had to try and prevent the memory of the object from entering awareness and maintain their attention on the cue word. If the object came to mind, they were asked to push it out of their mind and to report the intrusion at the end of the trial.

We used the Hierarchical Gaussian Filter (Mathys et al., 2011) to estimate participants' beliefs about upcoming intrusive memories and then we used the resulting trajectories of beliefs and prediction errors as parametric modulators of the top-down connectivity between hubs of the brain control system and the memory system. The control system included the anterior and posterior parts of the MFG and the memory control regions were composed by the rostral hippocampus, the caudal hippocampus and the precuneus. For both T1 and T2, the regions of interest (ROI) were built by averaging the time series of the 30 contiguous voxels of the peak of activity for each ROI (using no-think > think contrast for aMFG and pMFG, and no-think < think contrast for memory regions) within the pre-selected ROIs. Importantly, although this procedure results in different peak locations at T1 and T2, the constraints constituted by the pre-defined ROIs ensure the interpretability of longitudinal comparisons. Details can be found in Leone et al. (under review).

Hippocampal subfields

All the participants were scanned at both time points T1 and T2 with a 3T Achieva MRI scanner (Philips) at the Cyceron Center (Caen, France). A high-resolution proton density weighted sequence was also acquired perpendicularly to the long axis of the hippocampus (TR = 6500 ms; TE = 80 ms; flip angle = 90°; in-plane resolution = 0.391 × 0.391 mm^2; slice thickness = 2 mm; no gap; 30 slices) in order to segment HS. HS segmentations for both T1 and T2 have been implemented as in Postel et al. (2019). Briefly, hippocampal subfields were segmented with the software *ASHS* (Yushkevich et al., 2015), using an homemade atlas based on both trauma-exposed and non-exposed populations. Bilateral hippocampus was segmented into four different subfields: CA1, CA2-3/DG, Subiculum and Tail. We decided to include CA2, CA3 and DG in a unique region because of the absence of clear anatomical landmarks on MRI images and the limited sizes of these subfields. The automatic segmentations were visually checked before extraction of the volumes for statistical analyses. As we did not have any hypothesis on the lateralization of hippocampal dysfunctions in PTSD, we averaged left and right hemispheres for our analyses.

References

Admon, R., Leykin, D., Lubin, G., Engert, V., Andrews, J., Pruessner, J., & Hendler, T. (2013). Stress-induced reduction in hippocampal volume and connectivity with the ventromedial prefrontal cortex are related to maladaptive responses to stressful military service. *Human Brain Mapping*, *34*(11), 2808–2816. https://doi.org/10.1002/hbm.22100

Anderson, M. C., Bunce, J. G., & Barbas, H. (2016). Prefrontal-hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory*, *134 Pt A*, 145–161. https://doi.org/10.1016/j.nlm.2015.11.008

Apfel, B. A., Ross, J., Hlavin, J., Meyerhoff, D. J., Metzler, T. J., Marmar, C. R., Weiner, M. W., Schuff, N., & Neylan, T. C. (2011). Hippocampal volume differences in Gulf War veterans with current versus lifetime posttraumatic stress disorder symptoms. *Biological Psychiatry*, *69*(6), 541–548. https://doi.org/10.1016/j.biopsych.2010.09.044

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, *10*(6), 410–422. https://doi.org/10.1038/nrn2648

Barron, H. C., Vogels, T. P., Emir, U. E., Makin, T. R., O'Shea, J., Clare, S., Jbabdi, S., Dolan, R. J., & Behrens, T. E. J. (2016). Unmasking Latent Inhibitory Connections in Human Cortex to Reveal Dormant Cortical Memories. *Neuron*, *90*(1), 191–203. https://doi.org/10.1016/j.neuron.2016.02.031

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Bergmann, O., Spalding, K. L., & Frisén, J. (2015). Adult Neurogenesis in Humans. *Cold Spring Harbor Perspectives in Biology*, *7*(7), a018994. https://doi.org/10.1101/cshperspect.a018994

Besnard, A., & Sahay, A. (2016). Adult Hippocampal Neurogenesis, Fear Generalization, and Stress. *Neuropsychopharmacology*, *41*(1), 24–44. https://doi.org/10.1038/npp.2015.167

Bisby, J., & Burgess, N. (2017). Differential effects of negative emotion on memory for items and associations, and their relationship to intrusive imagery. *Current Opinion in Behavioral Sciences*, *17*, 124–132. https://doi.org/10.1016/j.cobeha.2017.07.012

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. *Journal of Traumatic Stress*, *28*(6), 489–498. https://doi.org/10.1002/jts.22059

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychological Review*, *117*(1), 210–232. https://doi.org/10.1037/a0018113

Butts, K. A., Weinberg, J., Young, A. H., & Phillips, A. G. (2011). Glucocorticoid receptors in the prefrontal cortex regulate stress-evoked dopamine efflux and aspects of executive function. *Proceedings of the National Academy of Sciences*, *108*(45), 18459–18464. https://doi.org/10.1073/pnas.1111746108

Carr, V. A., Rissman, J., & Wagner, A. D. (2010). Imaging the human medial temporal lobe with high-resolution fMRI. *Neuron*, *65*(3), 298–308. https://doi.org/10.1016/j.neuron.2009.12.022

Czéh, B., Varga, Z. K. K., Henningsen, K., Kovács, G. L., Miseta, A., & Wiborg, O. (2015). Chronic stress reduces the number of GABAergic interneurons in the adult rat hippocampus, dorsal-ventral and region-specific differences. *Hippocampus*, *25*(3), 393–405. https://doi.org/10.1002/hipo.22382

Daunizeau, J., Ouden, H. E. M. den, Pessiglione, M., Kiebel, S. J., Stephan, K. E., & Friston, K. J. (2010). Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. *PLOS ONE*, *5*(12), e15554. https://doi.org/10.1371/journal.pone.0015554

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Gagnon, S. A., & Wagner, A. D. (2016). Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences: Acute stress and episodic memory retrieval. *Annals of the New York Academy of Sciences*, *1369*(1), 55–75. https://doi.org/10.1111/nyas.12996

Gao, J., Wang, H., Liu, Y., Li, Y., Chen, C., Liu, L., Wu, Y., Li, S., & Yang, C. (2014). Glutamate and GABA imbalance promotes neuronal apoptosis in hippocampus after stress. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, *20*, 499–512. https://doi.org/10.12659/MSM.890589

Gilbertson, M. W., Shenton, M. E., Ciszewski, A., Kasai, K., Lasko, N. B., Orr, S. P., & Pitman, R. K. (2002). Smaller hippocampal volume predicts pathologic vulnerability to psychological trauma. *Nature Neuroscience*, *5*(11), 1242–1247. https://doi.org/10.1038/nn958

Gould, E. (2007). Structural plasticity. In P. Andersen, R. Morris, D. Amaral, T. Bliss, & J. O'Keefe (Eds.), *The hippocampus book* (pp. 321–341). Oxford University Press.

Gunn, B., Brown, A., Lambert, J., & Belelli, D. (2011). Neurosteroids and GABAA Receptor Interactions: A Focus on Stress. *Frontiers in Neuroscience*, *5*, 131. https://doi.org/10.3389/fnins.2011.00131

Harvey, B. H., Oosthuizen, F., Brand, L., Wegener, G., & Stein, D. J. (2004). Stress–restress evokes sustained iNOS activity and altered GABA levels and NMDA receptors in rat hippocampus. *Psychopharmacology*, *175*(4), 494–502. https://doi.org/10.1007/s00213-004-1836-4

Kaouane, N., Porte, Y., Vallée, M., Brayda-Bruno, L., Mons, N., Calandreau, L., Marighetto, A., Piazza, P. V., & Desmedt, A. (2012). Glucocorticoids Can Induce PTSD-Like Memory Impairments in Mice. *Science*, *335*(6075), 1510–1513. https://doi.org/10.1126/science.1207615

Kheirbek, M. A., Klemenhagen, K. C., Sahay, A., & Hen, R. (2012). Neurogenesis and generalization: A new approach to stratify and treat anxiety disorders. *Nature Neuroscience*, *15*(12), 1613–1620. https://doi.org/10.1038/nn.3262

Kida, S. (2019). Reconsolidation/destabilization, extinction and forgetting of fear memory as therapeutic targets for PTSD. *Psychopharmacology*, *236*(1), 49–57. https://doi.org/10.1007/s00213-018-5086-2

Koch, S. B. J., van Ast, V. A., Kaldewaij, R., Hashemi, M. M., Zhang, W., Klumpers, F., & Roelofs, K. (2021). Larger dentate gyrus volume as predisposing resilience factor for the development of trauma-related symptoms. *Neuropsychopharmacology*, *46*(7), 1283–1292. https://doi.org/10.1038/s41386-020-00947-7

Leone, G., Postel, C., Mary, A., Fraisse, F., Vallée, T., Viader, F., Peschanski, D., Dayan, J., Eustache, F., & Gagnepain, P. (under review). *Predicting the future to suppress the past after trauma*.

Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta Psychologica*, *127*(3), 623–635. https://doi.org/10.1016/j.actpsy.2007.12.004

Liberzon, I., & Abelson, J. L. (2016). Context Processing and the Neurobiology of Post-Traumatic Stress Disorder. *Neuron*, *92*(1), 14–30. https://doi.org/10.1016/j.neuron.2016.09.039

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., Vallée, T., Klein-Peschanski, C., Viader, F., Sayette, V. de la, Peschanski, D., Eustache, F., & Gagnepain, P. (2020). Resilience after trauma: The role of memory suppression. *Science*, *367*(6479). https://doi.org/10.1126/science.aay8477

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian Foundation for Individual Learning Under Uncertainty. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00039

McEwen, B. S., Nasca, C., & Gray, J. D. (2016a). Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. *Neuropsychopharmacology*, *41*(1), 3–23. https://doi.org/10.1038/npp.2015.171

McEwen, B. S., Nasca, C., & Gray, J. D. (2016b). Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex. *Neuropsychopharmacology*, *41*(1), 3–23. https://doi.org/10.1038/npp.2015.171

Moynihan, J. A., Chapman, B. P., Klorman, R., Krasner, M. S., Duberstein, P. R., Brown, K. W., & Talbot, N. L. (2013). Mindfulness-based stress reduction for older adults: Effects on executive function, frontal alpha asymmetry and immune function. *Neuropsychobiology*, *68*(1), 34–43. https://doi.org/10.1159/000350949

Popoli, M., Yan, Z., McEwen, B. S., & Sanacora, G. (2012). The stressed synapse: The impact of stress and glucocorticoids on glutamate transmission. *Nature Reviews Neuroscience*, *13*(1), 22–37. https://doi.org/10.1038/nrn3138

Postel, C., Mary, A., Dayan, J., Fraisse, F., Vallée, T., Guillery-Girard, B., Viader, F., Sayette, V. de la, Peschanski, D., Eustache, F., & Gagnepain, P. (2021). Variations in response to trauma and hippocampal subfield changes. *Neurobiology of Stress*, *15*, 100346. https://doi.org/10.1016/j.ynstr.2021.100346

Postel, C., Viard, A., André, C., Guénolé, F., de Flores, R., Baleyte, J., Gerardin, P., Eustache, F., Dayan, J., & Guillery-Girard, B. (2019). Hippocampal subfields alterations in adolescents with post-traumatic stress disorder. *Human Brain Mapping*, *40*(4), 1244–1252. https://doi.org/10.1002/hbm.24443

Qin, S., Hermans, E. J., van Marle, H. J. F., Luo, J., & Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological Psychiatry*, *66*(1), 25–32. https://doi.org/10.1016/j.biopsych.2009.03.006

Rolls, E. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*, 74. https://doi.org/10.3389/fnsys.2013.00074

Rolls, E. T. (2010). A computational theory of episodic memory formation in the hippocampus. *Behavioural Brain Research*, *215*(2), 180–196. https://doi.org/10.1016/j.bbr.2010.03.027

Schmitz, T. W., Correia, M. M., Ferreira, C. S., Prescot, A. P., & Anderson, M. C. (2017). Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nature Communications*, *8*(1), 1311. https://doi.org/10.1038/s41467-017-00956-z

Schoenfeld, T. J., McCausland, H. C., Morris, H. D., Padmanaban, V., & Cameron, H. A. (2017). Stress and Loss of Adult Neurogenesis Differentially Reduce Hippocampal Volume. *Biological Psychiatry*, *82*(12), 914–923. https://doi.org/10.1016/j.biopsych.2017.05.013

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of Human Memory: Brain Mechanisms and Clinical Relevance. *Biological Psychiatry*, *76*(4), 274–280. https://doi.org/10.1016/j.biopsych.2014.03.008

Shields, G. S., Sazma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, *68*, 651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

Sinclair, A. H., & Barense, M. D. (2019). Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends in Neurosciences*, *42*(10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Speckens, A. E. M., Ehlers, A., Hackmann, A., Ruths, F. A., & Clark, D. M. (2007). Intrusive memories and rumination in patients with post-traumatic stress disorder: A phenomenological comparison. *Memory*, *15*(3), 249–257. https://doi.org/10.1080/09658210701256449

Steiger, F., Nees, F., Wicking, M., Lang, S., & Flor, H. (2015). Behavioral and central correlates of contextual fear learning and contextual modulation of cued fear in posttraumatic stress disorder. *International Journal of Psychophysiology*, *98*(3), 584–593. https://doi.org/10.1016/j.ijpsycho.2015.06.009

Surget, A., & Belzung, C. (2021). Adult hippocampal neurogenesis shapes adaptation and improves stress response: A mechanistic and integrative perspective. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-021-01136-8

van der Kolk, B. A. (2007). The history of trauma in psychiatry. In *Handbook of PTSD: Science and practice* (pp. 19–36). The Guilford Press.

van Marle, H. (2015). PTSD as a memory disorder. *European Journal of Psychotraumatology*, *6*(1), 27633. https://doi.org/10.3402/ejpt.v6.27633

Weis, C. N., Webb, E. K., Huggins, A. A., Kallenbach, M., Miskovich, T. A., Fitzgerald, J. M., Bennett, K. P., Krukowski, J. L., deRoon-Cassini, T. A., & Larson, C. L. (2021). Stability of hippocampal subfield volumes after trauma and relationship to development of PTSD symptoms. *NeuroImage*, *236*, 118076. https://doi.org/10.1016/j.neuroimage.2021.118076

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yuen, E. Y., Wei, J., Liu, W., Zhong, P., Li, X., & Yan, Z. (2012). Repeated Stress Causes Cognitive Impairment by Suppressing Glutamate Receptor Expression and Function in Prefrontal Cortex. *Neuron*, *73*(5), 962–977. https://doi.org/10.1016/j.neuron.2011.12.033

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L., Kliot, D., Das, S. R., & Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, *36*(1), 258–287. https://doi.org/10.1002/hbm.22627

Zlotnick, C., Franklin, C. L., & Zimmerman, M. (2002). Does "subthreshold" posttraumatic stress disorder have any clinical relevance? *Comprehensive Psychiatry*, *43*(6), 413–419. https://doi.org/10.1053/comp.2002.35900

DISCUSSION

9. Synthesis of the main findings

In order to survive, living beings need to adapt to a world whose rules are often beyond their comprehension. The memory traces of the past experiences are essential for our interpretation of the world. An extremely harmful experience falling outside the range of expected possibilities can change the interpretation of reality. For instance, an event threatening a person's survival or physical integrity such as a terrorist attack may lead to the overestimation of the probabilities that other assaults to his/her life will occur in future. This new internal model of the world as a dangerous and threatening place may in turn lead the person to avoid any encounters with possible threats in order to maximize his/her chances of surviving. While, in the first instance, this may seem evolutionarily adaptive, people developing PTSD following a traumatic experience tend to **overgeneralize fears** to non-threatening environmental stimuli, often lacking of a meaningful direct link with the trauma.

In PTSD, the persistence of trauma-related, vivid, intrusive memories plays a key role in the maintenance of maladaptive avoidance, distress and negative emotions. Intrusive memories involve the transient formation of mental trauma-related perceptual traces, often accompanied by increased autonomic response and strong emotional reactivations, and reportedly lived similarly to experiencing the trauma again. Re-experiencing the trauma could also strengthen the associations between safe cues and threatening outcomes and favour avoidance in several ways. Intrusive memories are characterized by intense perceptual and emotional contents and are often triggered by the physical properties of environmental stimuli, detached from their contexts. Environmental stimuli sharing more or less conscious low-level properties with the content of the intrusive memories could be strongly associated with the trauma, leading to overgeneralization of fear. In turn, these aberrant associations could favour the avoidance of the stimuli learnt to trigger intrusive memories. In this vicious circle, avoiding the exposition to neutral stimuli that might remind the trauma would prevent the possibility to update these aberrant associations.

In the first study, we investigated whether **intrusive memories** could be the target of avoidant strategies. Despite, at a first glance, this may seem a paradox, as avoiding environmental stimuli can leave their associations with threatening outcomes unaltered,

avoiding intrusive memories could leave their memory trace unaltered. After having shown a general dysfunction in the brain network normally supporting the control of memory intrusion (Mary et al. 2020) in PTSD, in the first study of this thesis we hypothesized that this dysfunction may root on the dual nature of memory control mechanisms. On the one hand, predictive control mechanisms would be deployed to proactively prevent the insurgence of intrusive memories and, on the other hand, reactive control mechanisms would be added when intrusive memories enter consciousness. In a brain connectivity point of view, predictive control would adapt the amount of MFG-guided top-down inhibition over the hippocampus and the precuneus to the expected need for control (i.e., beliefs) and reactive mechanisms would respond to PE signals, in order to deploy the additional control required by the intrusive memories in the TNT task. As it has been previously shown that the memory control dysfunctions in PTSD are generalized to neutral stimuli and not specific to the traumatic memories (Mary et al. 2020), the materials of the TNT were neutral. The absence of trauma-related materials allows investigating the general mechanisms of memory control, ensuring to put the PTSD+ group and the non-PTSD groups in an equal footage.

We used computational modelling to infer beliefs of participants about the upcoming intrusive memories. We then modelled predictive and reactive memory control as the modulation of beliefs and PE, respectively, in the MFG top-down inhibition over the hippocampus and the precuneus using DCM. We found that nonexposed and resilient individuals harmoniously balanced predictive and reactive control, while individuals with PTSD did not. Participants who developed PTSD following to Paris November 2015 terrorist attacks proactively avoided intrusive memories, as shown by exaggerated predictive control when compared to resilient and nonexposed individuals. At the same time, we observed that individuals with PTSD were not able to apply PE-guided reactive control when intrusive memories needed to be purged away, as shown by the absence of PE modulation on the MFG coupling with the hippocampus in this group. Altogether, our results pointed out at the **imbalance between predictive and reactive control** as a specific risk factor for PTSD. Accordingly, the higher degrees of imbalance towards predictive control correlated with higher severity of intrusion and avoidance symptoms, but not with transdiagnostic clinical features.

A previous study of our research group, in the same cohort, has shown reductions in the hippocampal volumes in participants with PTSD. Specifically, PTSD+ showed reductions in CA and CA2-3/DG volumes when compared to nonexposed and resilient individuals, and

these reductions were correlated with intrusion and avoidance symptoms, respectively. In the second study of this thesis, we used a longitudinal design to address a long-lasting question: do hippocampal and memory control alterations are more likely to precede the trauma, constituting a pre-existing risk factor, or to follow the trauma as an effect of stress? To answer this important clinical question, we first investigated whether remission from PTSD four years after the trauma was related changes in memory control strategies and hippocampal plasticity.

We found that participants remitting from PTSD were characterized by the recovery of balance between predictive and reactive control and significant plastic changes in the left CA1 and the right CA2-3/DG volumes. On the contrary, participants with stable PTSD showed significant reductions in CA2-3/DG. When considering all the participants with PTSD two years after the trauma, we found that the recovery of memory control balance was specifically related to reductions in intrusive symptoms five years after the trauma. We also found that slight plastic changes of CA2-3/DG were predictive of future PTSD symptoms' general reductions. These results suggest that memory control and hippocampal disorders in PTSD may follow the trauma, as a possible effect of stress, and changes in memory control and hippocampal plasticity predict the future clinical evolution.

In the next sections, we integrate our results in the context of the theories proposing PTSD as a hippocampal disorder and the accounts of PTSD as a memory control disorders. While both accounts have up to now explained important aspects of PTSD, a link between the two is missing. We hope to furnish a unified model of PTSD incorporating hippocampal and memory control dysfunctions as a consequence of a maladaptive stress response.

10. Hippocampal model of PTSD

Our longitudinal results showed that hippocampal plasticity in time forecasted future reductions in avoidance severity, and recovery from PTSD was associated with increased left CA1 and right CA2-3/DG volumes in time. Although whether reduced hippocampal volume represents a pre-existing risk factor or an abnormal consequence of the exposure to intense stress is still debated, these results suggest that a maladaptive reaction to acute stress might have modified the hippocampal structure.

The hippocampal-amygdalar circuit is the core neurobiological substrate of the Brewin's dual representation model, detailed in the paragraph **PTSD as a memory disorder**. This model proposes that PTSD is characterized by the overconsolidation of emotional trauma-related memories and the parallel deficit in the contextual integration of the traumatic engrams (Brewin et al. 2010b). The original model assumes that the neural bases of these alterations rely on the hyperactivity of the amygdala and the hypoactivity of the hippocampus, respectively.

Stress can influence the hippocampal plasticity and memory consolidation. An intense stress cause the activation of both the sympathetic nervous system and the hypothalamic–pituitary–adrenal (HPA) axis. Stress typically enhance memory consolidation, and intense emotional memories are more likely to be encoded in the long term. In the immediate aftermath of a stressful experience, the sympathetic system releases glucocorticoids, which rapidly cross the blood-brain barrier, activating the limbic brain areas, specifically targeting the noradrenergic projection to the **basolateral amygdala** (BLA). The BLA then modulates the activity of the hippocampus and the prefrontal cortex (Schwabe et al. 2012). Animal studies have demonstrated that reducing **glucocorticoids** signalling impair memory consolidation.

Crucially, the effect of stress-related glucocorticoids on memory consolidation is not linear: several studies have demonstrated that a transient pharmaceutical augmentation of hippocampal glucocorticoids enhance memory, while the prolonged release of this hormone impair memory performances (Oitzl, Fluttert, and de Kloet 1998). The relationship between

stress and memory appears as an inversed-U: low concentrations of glucocorticoids would have a beneficial effect and high concentrations would have a detrimental effect.

An alternative model proposes that the effect of stress on hippocampal-dependent memory consolidation may be time-dependent. Accordingly, in the first 30 minutes following the exposure to an acute stressor, high levels of catecholamine and fast nongenomic glucocorticoid signalling would increase the activity of the BLA, which in turn, would impair the PFC functioning and enhance hippocampal synaptic plasticity via LTP. These biological processes would result in an increasing of memory storage and the parallel reduction of recall of memories unrelated to the stressor. After one-to-two hours, the normalization of catecholamine levels and the slow glucocorticoids modifications in gene transcriptions lead to the inhibition of the hippocampal plasticity, resulting in reduced storage and retrieval of the information unrelated to the stressor (see **Figure 13**). Importantly, this reduction in the hippocampal plasticity aims avoiding the interference of new memories and favour the long-term consolidation of the emotional memory (Gagnon and Wagner 2016).

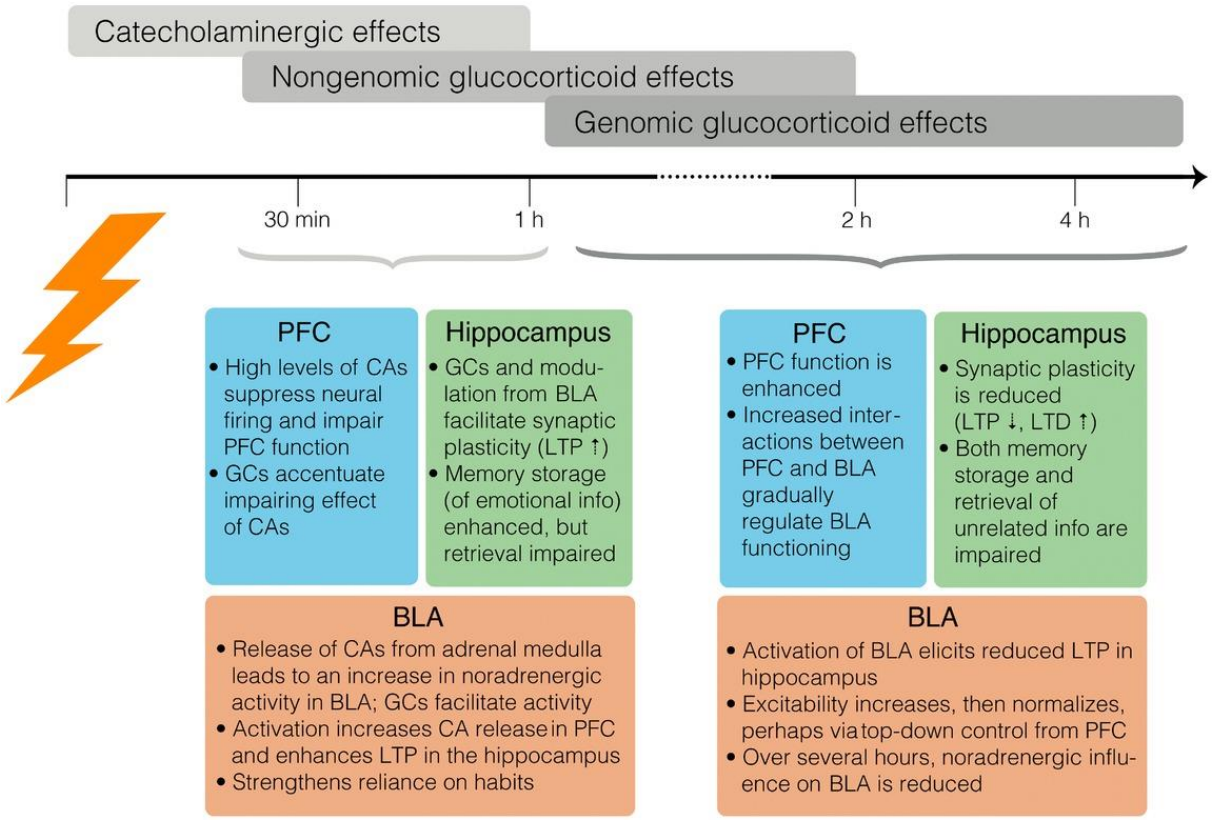


Figure 13. Schematic representation of the effect of stress on memory over time. Adapted from (Gagnon and Wagner 2016).

Chronic stress causing the prolonged exposure to glucocorticoids may alter structure and functioning of the hippocampus in the long period (Gagnon and Wagner 2016). Compatibly with this hypothesis, a study has shown that higher salivary levels of cortisol correlated with the reduced right hippocampal volumes in individuals with PTSD (Lindauer et al. 2006). However, the specific role of cortisol in PTSD is still not completely understood, with some human studies reporting increased cortisol and others reporting lower cortisol (see Pitman et al. 2012 for a review). On the contrary, evidence from animal models of PTSD have noticeably shown that pharmaceutically increasing hippocampal glucocorticoid receptors induced alterations in the hippocampal-amygdalar circuit (Kaouane et al. 2012).

Altogether, evidence that stress can enhance the long term encoding of memory traces supports the **emotional overconsolidation hypothesis**. Emotional memories, especially memories containing information fundamental for the individual's survival, are more salient and their storage is a priority over other types of memories, an effect known are flashbulb memories (Hirst and Phelps 2016). In individuals developing PTSD, an exaggerated cascade of glucocorticoids signalling consequent to the trauma may cause an exaggerated noradrenergical signalling of the BLA over the hippocampus, leading to the over-consolidation of the traumatic memory. However, it should be bear in mind that PTSD is characterized by the paradoxical presence of hypermnesia for sensorial implicit memories and amnesia for declarative memories of the trauma (Desmedt, Marighetto, and Piazza 2015). Under the emotional overconsolidation hypothesis, the amygdala would specifically elicit the overconsolidation of the sensorial and emotional aspects of the trauma, but not of the contextual details. In line with this hypothesis, it has been found that lesions in the amygdala impaired the encoding so-called "*gist memories"* (i.e., symbolic, sensorial and emotional representations), but not the encoding of the detailed memories (Adolphs, Tranel, and Buchanan 2005). Human fMRI studies have confirmed the influence of amygdala in memory encoding (Phelps 2004) and reported increased activity of this structure during the recall of the traumatic event (Layton and Krikorian 2002). In summary, an exaggerated amygdala response to the stress could lead to the strengthening of the sensory and emotional representations of the trauma

In PTSD, the emotional hypermnesia is accompanied by difficulties in the conscious recall of the contextual information about the trauma. Indeed, intrusive memories are transient intense sensory and emotional retrievals lacking of contextual information, likely due to amygdalar hyperactivity and the parallel hypoactivity of the hippocampus (Brewin et al.

2010a). This **deficit of contextual integration** would depend on the poor encoding of peritraumatic contextual information in the hippocampus. It has been shown that the hippocampus plays a specific role in the encoding and the retrieval of detailed contextual memories and, on the contrary, memories not containing contextual details can be retrieved without the involvement of this structure (Wiltgen et al. 2010). In PTSD, reductions in hippocampal neuronal density and integrity following intense and prolonged stress and the consequent deficits in hippocampal functioning could affect the possibility to recall contextual details (Garfinkel et al. 2014). Accordingly, a study using an animal model of PTSD has shown that increasing the concentration of the hippocampal glucocorticoids receptors affected the identification and learning of threatening contexts and provoked fear responses in safe environments (Kaouane et al. 2012). The deficits in contextual integration in PTSD are not necessarily trauma-specific, as shown by several studies reporting that individuals with PTSD have general difficulties in predicting safe and threatening contexts (Steiger et al. 2015).

The deficit in the contextual integration in PTSD may be rooted in the hippocampal functions underlying the associations between memory traces. Two distinct, yet complementary, properties in the hippocampal processing of associations between different stimuli should be distinguished. **Pattern completion** refers to the process allowing recovering a memory trace from a partial external cue. Given a new partial cue *A'*, the hippocampus is able to reactivate a complete stored representation *A* sharing elements with *A'* to fill the incomplete cue (see **Figure 14**, on the right). In parallel with the storage of similar patterns, memory retrieval requires the hippocampus to accurately separate distinct memory traces (Hunsaker and Kesner 2013). **Pattern separation** is the mechanisms allowing to dissociate two different memory traces *A* and *A'* sharing common elements (see **Figure 14**, on the left). While pattern completion aims responding to an incomplete input with a previously store pattern, pattern separation aims making similar inputs less similar (Guzowski, Knierim, and Moser 2004). Pattern completion and separation are two fundamental functions to integrate and discriminate similar experiences and are they constitute the bases of episodic memory. For a complete overview of these hippocampal processes, please see Rolls (2013) and Yassa and Stark (2011).

Different anatomical substrates of pattern completion and pattern separation have been found in the hippocampus. Hippocampal subfields are characterized by different histological properties and different functional specializations (Duvernoy 2005). **CA1** is preferentially involved in pattern completion. An early neurophysiology study has found that the CA1

pyramidal cells' firing was specifically responsible for pattern completion and the suppression of the CA3 output to CA1 did not impaired this process (Mizumori et al. 1989). Following studies have suggested that CA1 combines together the separated representations stored in CA3, creating the whole episodic representations by completing partially overlapping traces (Rolls 2010).

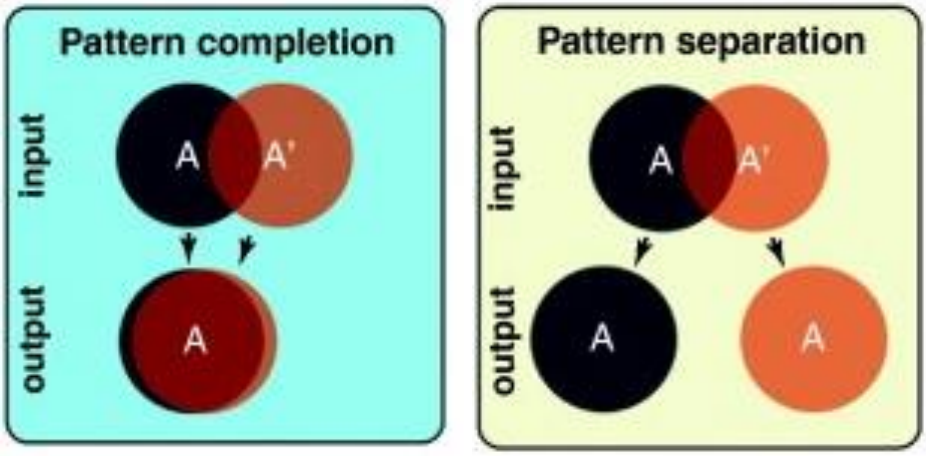


Figure 14. Pattern completion and pattern separation. Adapted from (Yassa and Stark 2011).

Reduced neuronal density and integrity in CA1 and CA3/DG in PTSD may compromise pattern completion and separation. Compatibly with this hypothesis, reductions in volumes of these hippocampal subfields have been reported in individuals developing PTSD following a traumatic experience but not in resilient individuals (Chen et al. 2018; Hayes et al. 2011; Postel et al. 2021). Notably, alterations in CA1 would alter pattern completion, possibly explaining why even environmental cues weakly associated with the trauma can trigger intrusive memories. Accordingly, in the study of (Postel et al. 2021), higher CA1 atrophy was correlated with higher intrusive symptoms' severity. Animal studies have been found that the stress-related atrophy of CA1 may be driven by the loss of GABAergic interneurons (Czéh et al. 2015; McEwen, Nasca, and Gray 2016). Altogether, these findings could suggest that intrusive memories may arise by CA1-mediated alterations in pattern completion, causing to the lack of inhibition of the traumatic memory traces when confronted with an environmental cue showing even weak common elements. In this context, the deficit in contextual integration would be reflect the lack of completion of the emotional and the peritraumatic elements of the traumatic memory.

In our longitudinal study (i.e., the second study of this thesis), we found plastic increasing in the left CA1 in individuals remitting from PTSD. We also found that changes in CA1 volume between one and three years from the trauma years in individual with persisting PTSD were predictive of subsequent changes in both intrusive and avoidance symptoms. These results may suggest that changes in CA1 volumes, perhaps driven by the recovery of the GABAergic neuronal integrity, might facilitate of the excitation/inhibition balance in the hippocampus and predict slow changes in PTSD symptoms' severity, promoting the remission from PTSD.

CA3 and the **DG** have been reportedly shown to be preferentially involved in pattern separation. Particularly, CA3 granule cells show a specific firing rate for large sensory input changes, determining the separation of the two traces, which will be stored separately (Yassa and Stark 2011). The DG is a hippocampal subfield presenting a fundamental characteristic: this part of the hippocampus is the only one where neurogenesis is still possible at the adult age (Bergmann, Spalding, and Frisén 2015). Many factors can influence the proliferation and differentiation of granulate neurons in the DG, such as hormones, network activity and epigenetic modulations. The possibility to develop new neurons in the adult age promotes the contextual discrimination between different overlapping experiences and memories, enabling pattern separation (Surget and Belzung 2021). Indeed, neurogenesis could facilitate pattern separation by providing new granulate neurons able to create connections with neurons in CA3 to facilitate the encoding and storage of separated memory traces (Rolls 2010).

Reductions in CA3 and DG have been reported in PTSD (Postel et al. 2021; Wang et al. 2010). Reduced volumes in CA3 and the DG and the consequent impairment of pattern separation could prevent the possibility to form conjoint integrations of different objects in their own context. In PTSD, these deficits in pattern separation can compromise the ability to discriminate between safe and threatening environments, leading to overgeneralization of fears to non-threatening stimuli (Besnard and Sahay 2016). In turn, overgeneralization of fear would lead individuals with PTSD to avoid stimuli that might have even a weak association associated with the trauma. Compatibly with this hypothesis, a previous study has found a correlation between reductions of CA2-3/DG volumes and higher severity of avoidance symptoms (Postel et al. 2021).

In line with the importance of CA3 and DG in PTSD, in the second study of this thesis we reported a specific increase of the left CA2-3/DG in individuals remitting from PTSD and

decrease of the right CA2-3/DG in participants who continued to be diagnosed as PTSD. Beyond the lateralization of the plastic changes, on which we did not have hypotheses, these results suggest that the plasticity of this hippocampal subfield is a marker of remission. Despite we were not able to differentiate CA2-3 from DG, due to the absence of anatomical landmarks, this volumetric changes could be at least in part due to increased neurogenesis in the DG. Furthermore, when exploring individuals with persistent PTSD after three years from the trauma, we found that plastic changes in this hippocampal subfield correlated with future changes in avoidance symptoms' severity. These results show that increased neurogenesis in the hippocampus have a causal relationship with the subsequent reduction in symptoms' severity. However, the directionality of this relationship remains unknown.

Altogether, this evidence suggests that a hippocampal model of PTSD can include stress as a determining factor of hippocampal structural and functional impairments. In PTSD, an aberrant CA1-mediated pattern completion and impaired CA3-DG-mediated pattern separation could result in the loss of contextual integration of the traumatic memories, facilitating intrusive memories and avoidant behaviour. Crucially, we reported that the degree of hippocampal plasticity was related with remission from PTSD and symptoms attenuation, thereby suggesting that the recovery of hippocampal functioning, including neurogenesis, could regulate the concomitant and future recovery from PTSD symptoms.

11. Memory control model of PTSD

Intrusive memories are a central clinical feature of PTSD. Along with a memory disorder, PTSD may be defined as an active forgetting disorder. Little is known about the functioning of the brain mechanisms supporting the suppression of intrusive memories in PTSD. Only a few studies have so far addressed this question, suggesting a general dysfunction in the prefrontal control over the hippocampus during the suppression of intrusive memories (see paragraph **PTSD as an active forgetting disorder** and the **ANNEX**). A popular model assumes a dualism in the brain control mechanisms (Braver 2012). Accordingly, cognitive control can be reached via two distinct, yet complementary, mechanisms driven by different computational signals: proactive and reactive control. P**redictive control** (or proactive control) is a form of early control requiring the maintenance of the goal-relevant information in a sustained way. Predictive control biases attention towards the goals, reducing the available resources to process stimuli unrelated to the internal objectives (Braver 2012). This form of control requires a sustained attention and it is resource consuming. **Reactive control** is a transient form of late correction following the failure of predictive control. Reactive mechanisms are transiently deployed as a form of adjustment when the outcome of proactive control is different than expected.

Predictive and reactive control mechanisms have recently received interest in numerous fields of cognitive neuroscience, including response inhibition, cognitive flexibility and conflict-control (for an exhaustive review, see Jiang, Heller, and Egner, 2014). The brain attempts to produce an efficient model in order to predict the future states of the world and adapt its behavior to the expected environmental demands. Computationally, predictive control is guided by the beliefs (or predictions) about the amount of control required by the task. Reactive control would is guided by the divergence between the expected and the real amount of control to apply, that is, PE.

Bayesian models of the behaviour represent a unique tool for modelling internal, unobservable beliefs (see paragraph **Bayesian modelling of human behaviour**) and this

technique has been applied to inhibitory control in several studies. In an interesting study, Ide et al. (2013) estimated subjects' beliefs while performing a stop-signal task. The authors modelled the beliefs probability of encountering a stop trial and they found that higher expectations correlated with increased response times, indicating enhanced in predictive control. Other studies have confirmed the potential of Bayesian modelling in estimating beliefs driving predictive control in the motor and cognitive control research fields (Barceló 2021; Hu et al. 2016; Pezzulo and Ognibene 2012), increasing our knowledge on how human beings can predict and flexibly adapt to the environmental demands.

However, predictive and reactive control mechanisms have never been investigated before in the framework of memory suppression. Although the existence of these two different mechanisms has been proposed by an influential theoretical model (Anderson et al. 2016), no studies have attempted to explore these dynamics through computational modelling. Memory suppression engages the right dlPFC to reduce the activity of the hippocampus. According to the neurobiological model proposed by Anderson and colleagues, the dlPFC may orchestrate two different types of control of intrusive memories intervening at different temporal stages via the following distinct anatomical pathways.

- An early and sustained control would *proactively* target the inhibition of the entorhinal forwards and backwards inputs to the hippocampus. This form of control would act as an **entorhinal gate** would prevent the retrieval of an unwanted memory before it begins. The quiescence of the hippocampus observed during memory suppression would be the result of detaching the hippocampus from its inputs and outputs.

- A form of reactive control intervenes when the entorhinal gate fails and the unwanted memories enter consciousness. Through these *reactive* mechanisms, the dlPFC would directly inhibit the hippocampal activity via the **modulation of the thalamic projections** to the hippocampus, particularly the ones forming into the nucleus reuniens. The advantage of directly inhibiting the hippocampus would be the consequent disruption of the pattern completion processes triggered by the external reminder, resulting in the interruption of the retrieval process.

Previous studies have defined reactive control only basing on the presence of intrusive memories, without considering the confounding effect of the concomitant predictive control

(Levy and Anderson 2012b). The lack of a fine-grained definition of predictive and reactive control could have led to the misinterpretation of the brain connectivity observed during intrusive trials. Our combination of Bayesian modeling and DCM overcomes these limitations, allowing isolating beliefs-driven predictive and PE-driven reactive control of intrusive memories. For the first time, we were able to test empirically the hypothesis that memory control is reached via these two distinct mechanisms. We built a family of DCM models incorporating beliefs and PE modulation of the MFG-guided downregulation of the hippocampus and the precuneus. Crucially, we tested our main hypothesis against two alternative hypotheses, by building other two DCM families assuming respectively: 1) bottom-up connectivity between MFG and memory regions; 2) top-down inhibition not guided by beliefs and PE signals. Evidence demonstrated that our hypothesis was the most likely, above and beyond chance. Indeed, BMS analyses revealed that the models incorporating belief-driven predictive and PE-driven reactive control mechanisms were more likely to have generated the real fMRI activation data. Despite we did not test the hypothesis that different pathways are involved in these two control modalities, our DCM analyses do not contradict this idea. Indeed, DCM relies on the modeling of the causal connectivity between a set of ROIs, without any inference on the eventual anatomical pathways linking these regions.

Our neurocomputational model revealed that the origin of the memory control dysfunctions in PTSD previously observed in Mary et al. (2020) could rely in the disruption of the balance between predictive and reactive mechanisms towards the formers. Individuals with PTSD were characterized by exaggerated predictive control and inexistent reactive control. The exaggerated predictive control of intrusive memories could be maladaptive for several reasons. Accordingly, the avoidance of stressful intrusive memories and thoughts have been reported to correlate with increased trauma-related intrusions in PTSD (Harvey and Bryant 1998) and greater symptoms severity in patients with major depression (Brewin, Reynolds, and Tata 1999). Disproportionate efforts in the prevention of intrusive memories could mimic pathological mechanisms such as avoiding places, people or situations that might remind the trauma. The predictive avoidance of intrusive memories implies the persistence of to-be-avoided memory representations in mind, resulting in a paradoxical rebound, perhaps preventing forgetting.

As detailed above, the hippocampus plays a fundamental role in the integration of memory representations in their contexts. The sustained predictive gating of the hippocampal activity could impair the retrieval of the **contextual information** of the avoided memory,

rather than the memory trace itself. Accordingly, a study has found that a sustained suppression of hippocampal activity during memory control triggered memory loss for the events surrounding the suppressed item, but for the item itself (Hulbert, Henson, and Anderson 2016). This mechanism may be crucial to understand the maladaptive role of predictive control in PTSD. An excessive predictive hippocampal suppression would indeed cut off the hippocampus from its inputs and outputs, reducing its activity and facilitating the loss of the contextual information surrounding the traumatic memory, but leaving the emotional and sensorial features unaltered. This loss of contextual information would impair the ability to distinguish safe and harmful contexts associated with the trauma, contributing to the avoidance of the trauma reminders. Our results showing that the grater imbalance towards predictive control correlated with higher avoidance severity support this hypothesis.

Alternately, predictive control may prevent forgetting because the gating of the hippocampal activity would prevent the activation of the to-be-forgotten memory trace, leaving the engram unaltered and less susceptible to modifications. As introduced in the paragraph **How memories are stored**, memories become vulnerable to modifications during retrieval. By gating retrieval, predictive control would not alter the memory engrams, which would only be silenced and susceptible to future recall. Our results showing that imbalanced memory control towards a predictive mode in PTSD correlated with increased severity of intrusive symptoms suggest may suggest that the temporary memory traces would favor further intrusive memories.

On the contrary, reactive control increases the vulnerability of memory traces, facilitating forgetting. Forgetting is an act of memory reconsolidation. Retrieving a memory entails the trace entering in a vulnerable state (Kida, 2019; Schwabe et al., 2014), and the reactivation of the memory may be a necessary condition for forgetting. This idea is in line with several findings showing that the suppression of the hippocampal activity during intrusive trials causes forgetting but not during nonintrusive trials (Levy and Anderson 2012a). According to a recent hypothesis, the reconsolidation and, consequently, the destabilization, of a memory trace depend on the intensity of its reactivation (Sinclair and Barense 2019). According to this hypothesis, in a U-shaped relationship, intense memory activations would strengthen the memory trace and moderate memory activations would weaken the memory traces. The intrusive memories successfully controlled via reactive control are transitory moderate reactivations of the unwanted memory engram. Thus, these memories could potentially facilitate the destabilization and weakening of the unwanted

memory trace. A recent alternative hypothesis proposed the existence of “inhibitory engrams” that parallel the neuronal connections forming memories, silencing the activation of these excitatory engrams (Barron et al., 2016). Reactive control of intrusive memories could target the hippocampal GABAergic interneurons potentiating the connections of these inhibitory engrams, silencing the specific neuronal networks activating the unwanted memories.

The recovered reactive control was a marker of remission from PTSD and specifically associated with the future reduction of intrusive symptoms’ severity. These findings suggest that changes in predictive and reactive control may follow the trauma as a maladaptive response to stress and inaugurate PTSD symptoms. Stress can significantly impair executive function, including control (Arnsten 2009). An intense stress induce the relocation of the executive resources normally supporting inhibition, working memory and flexibility towards the handling of the stressor (Shields, Sazma, and Yonelinas 2016). At the same time, the release of glucocorticoids associated with the stress response can cause an increased glutamatergic signaling in the PFC (Popoli et al. 2012), which have detrimental effects on the PFC-dependent cognitive functions (Qin et al. 2009; Yuen et al. 2012). Evidence has shown that blocking glucocorticoid receptors in the PFC improved prefrontal executive functions (Butts et al. 2011), and reductions in stress could explain the recovered balance in memory control mechanisms in remitted PTSD.

12. Towards a unified model of PTSD

We have described a potential model of PTSD as a memory control disorder and a potential model of PTSD as a hippocampal disorder. These two accounts of PTSD focus on two apparently separate aspects of the relationship between the prefrontal cortex and the hippocampus. If we consider memory control as the top-down signal transmission from the prefrontal cortex to the hippocampus, a model of PTSD as a mere control disorder would only focus on the source and the transmission of this signalling, neglecting its target (i.e., the hippocampus). On the contrary, a model of PTSD as a mere hippocampal disorder would only focus on the terminal part of the prefrontal-hippocampal circuit. However, we have shown the potential of both control and hippocampal dysfunctions in distinguishing and predicting pathological and resilient outcomes following a traumatic experience, suggesting that these two disorders coexist. Our results suggest that PTSD may be rooted in a general dysfunction of the prefrontal-hippocampal network. However, the nature of the relationship between control and hippocampal dysfunctions is still unexplored. Different hypotheses can be formulated.

One hypothesis is that a combination of pre-existing and post-traumatic alterations in the hippocampus facilitate the persistence of intrusive memories, leading to stress-related alterations in the PFC. According to this hypothesis, alterations in the CA1 integrity would produce an aberrant pattern completion, causing external cues seemingly unrelated to the trauma to trigger intrusive memories. The frequent re-experiencing of the traumatic experience can contribute to the maintenance of high stress level. Reductions in neurogenesis and CA2-3/DG volumes may disrupt pattern separation, impairing a net separation between the past and the present, provoking the sensation of living the trauma again during intrusive memories. This extremely vivid re-experiencing of the trauma accompanied by exaggerated autonomic response and the sustained activation of the HPA axis would lead to cortisol-mediated negative effect on the functioning of the PFC. In a vicious circle, the stress-affected PFC would reduce its effectiveness in control intrusive memories (see Figure 15a).

A second hypothesis is that intrusive memories arise from the failure of the control mechanisms and the stress generated by their persistence would affect hippocampal integrity, via the mechanisms described above. Accordingly, the acute stress due to the traumatic experience would directly cause functional alterations in the PFC, impairing it effectiveness in controlling intrusive memories. The stress caused by re-experiencing would then induce structural changes in the hippocampus. However, this hypothesis claims a general dysfunction of the prefrontal cortex in PTSD. Contradicting this assumption, our neurocomputational model has shown that predictive and reactive dynamics underlie the suppression of intrusive memories, and only reactive mechanisms are impaired in PTSD. On the contrary, individuals with PTSD showed exaggerated predictive control.

Thus, a third hypothesis is that the prefrontal cortex orchestrates memory control basing on the wrong signal in PTSD. As detailed above, learning and prediction deficits have been described in PTSD (see the paragraph **PTSD as a prediction disorder**). The hippocampus has been reportedly shown to be involved in the computational processes underlying the generation of beliefs and the processing of PE (Den Ouden et al., 2012; Schapiro et al., 2012). The loss of hippocampal integrity may provoke aberrant beliefs about upcoming intrusions, which would lead to an excess of predictive control. Partially supporting this hypothesis, we found that individuals with PTSD were less prone to update their beliefs about particular items when they failed to suppress the associated intrusions. The crystallization of high expectations about upcoming intrusions might be linked to the exaggerated predictive control. In parallel, the hippocampus might present deficits in the processing of PE and the communication of the error signal to PFC may be compromised, resulting in the lack of reactive control. Further studies should investigate the backward PE signalling from the hippocampus to the PFC, in order to investigate whether the signal is lost before reaching the PFC or, alternatively, the signal is aberrantly processed by the PFC (see **Figure 15b**).

A fourth hypothesis is that a maladaptive response to stress affects the hippocampal integrity and in PTSD the loss of reactive control may be due to the loss of hippocampal GABAergic interneurons. It has been proposed that predictive memory control aims gating the hippocampus by cutting of its inputs and outputs, and reactive control directly target the inhibition of the hippocampal activity (Anderson et al., 2016). The suppression of unwanted memory has been proposed to depend on the availability of GABA in the hippocampus (Schmitz et al., 2017). In PTSD, stress-induced reductions in CA1 GABAergic interneurons

(Czéh et al., 2015) could specifically impair the efficacy of PFC direct inhibition over the hippocampus. However, while this hypothesis would explain the lack of reactive control, it does not tell anything about the augmented predictive control. An integration between the third and the fourth hypotheses would explain the imbalance between predictive in reactive control in PTSD. Accordingly, aberrant hippocampal predictions would lead to exaggerated predictive control and the disruption of the GABAergic hippocampal substrate would disrupt reactive control. This proposal has strong neurobiological motivations. However, little is known about hippocampal GABA in PTSD and its relationship with memory suppression. Further studies should combine computational modelling of predictive and reactive control of intrusive memories with in vivo measurements of GABA receptors' density and functioning in PTSD (see **Figure 15c**).

Beyond the possible interpretations of the relationship between memory control and hippocampal disorders, the two studies of this thesis have shown that the two disorders disambiguated different clinical outcomes in the presence of comparable traumatic experiences. We identified neurobiological markers of remission from PTSD in the recovery of the neurocognitive functions supporting memory and forgetting, suggesting that overcoming these dysfunctions can promote resilience. Furthermore, we reported the significant value of these markers in forecasting different clinical trajectories following a traumatic experience.

These results shed light on novel potential targets for the treatment of PTSD. Most of the current psychological treatments of PTSD involve, to some degree, the re-exposure to the trauma (Brewin, 2018). The exposition to sensible trauma-related material could be problematic in the clinical settings, especially with certain patients. We have shown that the dysfunctions in controlling intrusive memories are general and not confined to the traumatic experience. The alterations of the hippocampal contextual integration have been proposed to be generic as well (Steiger et al., 2015). According to our findings, the goal of an effective psychological treatment of PTSD should be:

- The overcoming of the avoidance of intrusive memories;
- The restoration of the ability to disengage from unwanted memories entering consciousness;
- The promotion of the neurogenesis in the dentate gyrus;
- The restoration of pattern separation and pattern completion.

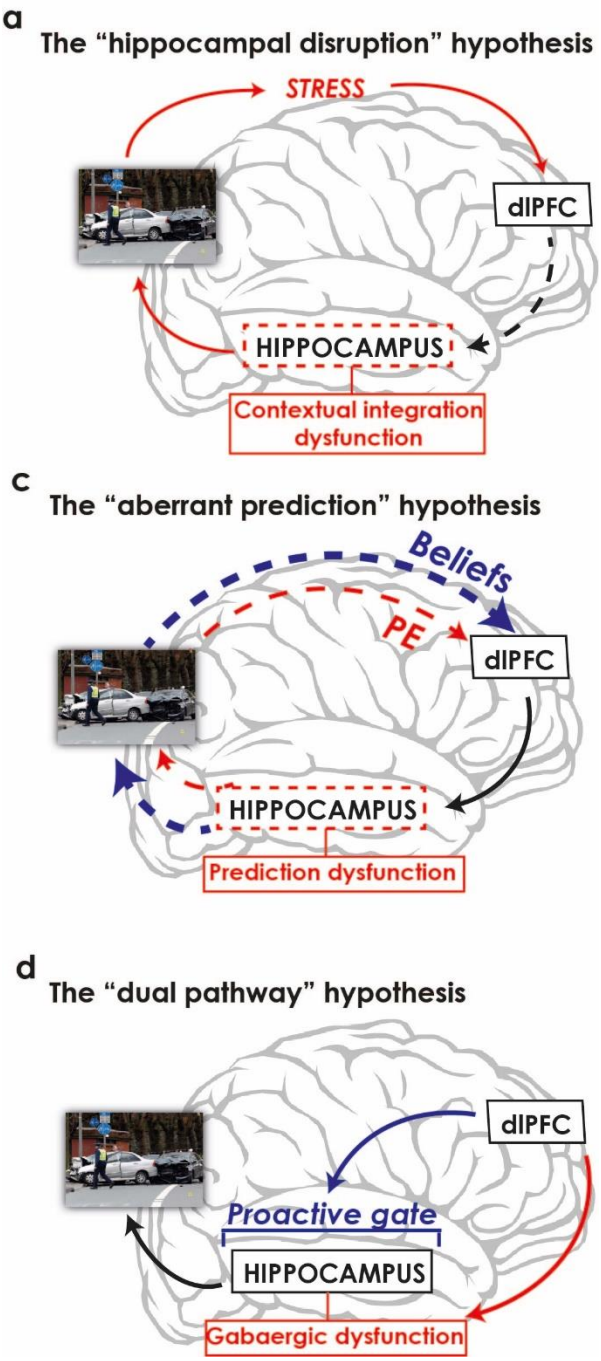


Figure 15. Different hypotheses on the relationship between memory control and hippocampal disorders in PTSD. Dashed lines indicate the origin of the disruption.

REFERENCES

Adams, Rick A., Quentin J. M. Huys, and Jonathan P. Roiser. 2016. 'Computational Psychiatry: Towards a Mathematically Informed Understanding of Mental Illness'. *Journal of Neurology, Neurosurgery & Psychiatry* 87(1):53–63. doi: 10.1136/jnnp-2015-310737.

Adolphs, Ralph, Daniel Tranel, and Tony W. Buchanan. 2005. 'Amygdala Damage Impairs Emotional Memory for Gist but Not Details of Complex Stimuli'. *Nature Neuroscience* 8(4):512–18. doi: 10.1038/nn1413.

American Psychiatric Association. 1952. *Diagnostic and Statistical Manual of Mental Disorders*. First Edition. American Psychiatric Association.

American Psychiatric Association. 1980. *Diagnostic and Statistical Manual of Mental Disorders*. Third Edition. American Psychiatric Association.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition. American Psychiatric Association.

Anderson, M. C., R. A. Bjork, and E. L. Bjork. 1994. 'Remembering Can Cause Forgetting: Retrieval Dynamics in Long-Term Memory'. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 20(5):1063–87. doi: 10.1037//0278-7393.20.5.1063.

Anderson, Michael C. 2003. 'Rethinking Interference Theory: Executive Control and the Mechanisms of Forgetting'. *Journal of Memory and Language* 49(4):415–45. doi: 10.1016/j.jml.2003.08.006.

Anderson, Michael C., Jamie G. Bunce, and Helen Barbas. 2016. 'Prefrontal-Hippocampal Pathways Underlying Inhibitory Control over Memory'. *Neurobiology of Learning and Memory* 134 Pt A:145–61. doi: 10.1016/j.nlm.2015.11.008.

Anderson, Michael C., and Collin Green. 2001. 'Suppressing Unwanted Memories by Executive Control'. *Nature* 410(6826):366–69. doi: 10.1038/35066572.

Anderson, Michael C., and Ean Huddleston. 2012. 'Towards a Cognitive and Neurobiological Model of Motivated Forgetting'. Pp. 53–120 in *True and False Recovered Memories*. Vol. 58, edited by R. F. Belli. New York, NY: Springer New York.

Anderson, Michael C., and Justin C. Hulbert. 2021. ‘Active Forgetting: Adaptation of Memory by Prefrontal Control’. *Annual Review of Psychology* 72(1):1–36. doi: 10.1146/annurev-psych-072720-094140.

Anderson, Michael C., Kevin N. Ochsner, Brice Kuhl, Jeffrey Cooper, Elaine Robertson, Susan W. Gabrieli, Gary H. Glover, and John D. E. Gabrieli. 2004. ‘Neural Systems Underlying the Suppression of Unwanted Memories’. *Science* 303(5655):232–35. doi: 10.1126/science.1089504.

Andreasen, Nancy C. 2010. ‘Posttraumatic Stress Disorder: A History and a Critique’. *Annals of the New York Academy of Sciences* 1208:67–71. doi: 10.1111/j.1749-6632.2010.05699.x.

Arnsten, Amy F. T. 2009. ‘Stress Signalling Pathways That Impair Prefrontal Cortex Structure and Function’. *Nature Reviews Neuroscience* 10(6):410–22. doi: 10.1038/nrn2648.

Aupperle, Robin L., Andrew J. Melrose, Murray B. Stein, and Martin P. Paulus. 2012. ‘Executive Function and PTSD: Disengaging from Trauma’. *Neuropharmacology* 62(2):686–94. doi: 10.1016/j.neuropharm.2011.02.008.

Badura-Brack, Amy, Timothy J. McDermott, Elizabeth Heinrichs-Graham, Tara J. Ryan, Maya M. Khanna, Daniel S. Pine, Yair Bar-Haim, and Tony W. Wilson. 2018. ‘Veterans with PTSD Demonstrate Amygdala Hyperactivity While Viewing Threatening Faces: A MEG Study’. *Biological Psychology* 132:228–32. doi: 10.1016/j.biopsycho.2018.01.005.

Banich, Marie T., Kristen L. Mackiewicz, Brendan E. Depue, Anson J. Whitmer, Gregory A. Miller, and Wendy Heller. 2009. ‘Cognitive Control Mechanisms, Emotion and Memory: A Neural Perspective with Implications for Psychopathology’. *Neuroscience & Biobehavioral Reviews* 33(5):613–30. doi: 10.1016/j.neubiorev.2008.09.010.

Barceló, Francisco. 2021. ‘A Predictive Processing Account of Card Sorting: Fast Proactive and Reactive Frontoparietal Cortical Dynamics during Inference and Learning of Perceptual Categories’. *Journal of Cognitive Neuroscience* 33(9):1636–56. doi: 10.1162/jocn_a_01662.

Barrientos, Sebastian A., and Vicente Tiznado. 2016. ‘Hippocampal CA1 Subregion as a Context Decoder’. *Journal of Neuroscience* 36(25):6602–4. doi: 10.1523/JNEUROSCI.1107-16.2016.

Benoit, Roland G., and Michael C. Anderson. 2012. ‘Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories’. *Neuron* 76(2):450–60. doi: 10.1016/j.neuron.2012.07.025.

Benoit, Roland G., Justin C. Hulbert, Ean Huddleston, and Michael C. Anderson. 2015. ‘Adaptive Top–Down Suppression of Hippocampal Activity and the Purging of Intrusive Memories from Consciousness’. *Journal of Cognitive Neuroscience* 27(1):96–111. doi: 10.1162/jocn_a_00696.

Bergmann, Olaf, Kirsty L. Spalding, and Jonas Frisén. 2015. ‘Adult Neurogenesis in Humans’. *Cold Spring Harbor Perspectives in Biology* 7(7):a018994. doi: 10.1101/cshperspect.a018994.

de Berker, Archy O., Robb B. Rutledge, Christoph Mathys, Louise Marshall, Gemma F. Cross, Raymond J. Dolan, and Sven Bestmann. 2016. ‘Computations of Uncertainty Mediate Acute Stress Responses in Humans’. *Nature Communications* 7(1):10996. doi: 10.1038/ncomms10996.

Besnard, Antoine, and Amar Sahay. 2016. ‘Adult Hippocampal Neurogenesis, Fear Generalization, and Stress’. *Neuropsychopharmacology* 41(1):24–44. doi: 10.1038/npp.2015.167.

Braver, Todd S. 2012. ‘The Variable Nature of Cognitive Control: A Dual Mechanisms Framework’. *Trends in Cognitive Sciences* 16(2):106–13. doi: 10.1016/j.tics.2011.12.010.

Brewin, C. R., T. Dalgleish, and S. Joseph. 1996. ‘A Dual Representation Theory of Posttraumatic Stress Disorder’. *Psychological Review* 103(4):670–86. doi: 10.1037/0033-295x.103.4.670.

Brewin, Chris R. 2011. ‘The Nature and Significance of Memory Disturbance in Posttraumatic Stress Disorder’. *Annual Review of Clinical Psychology* 7(1):203–27. doi: 10.1146/annurev-clinpsy-032210-104544.

Brewin, Chris R., James D. Gregory, Michelle Lipton, and Neil Burgess. 2010a. 'Intrusive Images in Psychological Disorders: Characteristics, Neural Mechanisms, and Treatment Implications.' *Psychological Review* 117(1):210–32. doi: 10.1037/a0018113.

Brewin, Chris R., James D. Gregory, Michelle Lipton, and Neil Burgess. 2010b. 'Intrusive Images in Psychological Disorders: Characteristics, Neural Mechanisms, and Treatment Implications'. *Psychological Review* 117(1):210–32. doi: 10.1037/a0018113.

Brewin, Chris R., Martina Reynolds, and Philip Tata. 1999. 'Autobiographical Memory Processes and the Course of Depression'. *Journal of Abnormal Psychology* 108(3):511–17. doi: 10.1037/0021-843X.108.3.511.

Brown, John. 1958. 'Some Tests of the Decay Theory of Immediate Memory'. *Quarterly Journal of Experimental Psychology* 10(1):12–21. doi: 10.1080/17470215808416249.

Brown, Vanessa M., Lusha Zhu, John M. Wang, B. Christopher Frueh, Brooks King-Casas, and Pearl H. Chiu. 2018. 'Associability-Modulated Loss Learning Is Increased in Posttraumatic Stress Disorder' edited by M. J. Frank. *ELife* 7:e30150. doi: 10.7554/eLife.30150.

Bryant, Richard A., Mark Creamer, Meaghan O'Donnell, David Forbes, Alexander C. McFarlane, Derrick Silove, and Dusan Hadzi-Pavlovic. 2017. 'Acute and Chronic Posttraumatic Stress Symptoms in the Emergence of Posttraumatic Stress Disorder: A Network Analysis'. *JAMA Psychiatry* 74(2):135–42. doi: 10.1001/jamapsychiatry.2016.3470.

Butts, Kelly A., Joanne Weinberg, Allan H. Young, and Anthony G. Phillips. 2011. 'Glucocorticoid Receptors in the Prefrontal Cortex Regulate Stress-Evoked Dopamine Efflux and Aspects of Executive Function'. *Proceedings of the National Academy of Sciences* 108(45):18459–64. doi: 10.1073/pnas.1111746108.

Catarino, Ana, Charlotte S. Küpper, Aliza Werner-Seidler, Tim Dalgleish, and Michael C. Anderson. 2015. 'Failing to Forget: Inhibitory-Control Deficits Compromise Memory Suppression in Posttraumatic Stress Disorder'. *Psychological Science* 26(5):604–16. doi: 10.1177/0956797615569889.

Chen, Lyon W., Delin Sun, Sarah L. Davis, Courtney C. Haswell, Emily L. Dennis, Chelsea A. Swanson, Christopher D. Whelan, Boris Gutman, Neda Jahanshad, Juan Eugenio Iglesias, Paul Thompson, Mid-Atlantic MIRECC Workgroup, H. Ryan Wagner, Philipp Saemann, Kevin S. LaBar, and Rajendra A. Morey. 2018. 'Smaller Hippocampal CA1 Subfield Volume in Posttraumatic Stress Disorder'. *Depression and Anxiety* 35(11):1018–29. doi: 10.1002/da.22833.

Corlett, Philip R., and Paul C. Fletcher. 2014. 'Computational Psychiatry: A Rosetta Stone Linking the Brain to Mental Illness'. *The Lancet. Psychiatry* 1(5):399–402. doi: 10.1016/S2215-0366(14)70298-6.

Cowan, Nelson. 2008. 'What Are the Differences between Long-Term, Short-Term, and Working Memory?' *Progress in Brain Research* 169:323–38. doi: 10.1016/S0079-6123(07)00020-9.

Crocq, Marc-Antoine, and Louis Crocq. 2000. 'From Shell Shock and War Neurosis to Posttraumatic Stress Disorder: A History of Psychotraumatology'. *Dialogues in Clinical Neuroscience* 2(1):47–55.

Czéh, Boldizsár, Zsófia K. Kalangyáné Varga, Kim Henningsen, Gábor L. Kovács, Attila Miseta, and Ove Wiborg. 2015. 'Chronic Stress Reduces the Number of GABAergic Interneurons in the Adult Rat Hippocampus, Dorsal-Ventral and Region-Specific Differences'. *Hippocampus* 25(3):393–405. doi: 10.1002/hipo.22382.

Daunizeau, Jean, Hanneke E. M. den Ouden, Matthias Pessiglione, Stefan J. Kiebel, Karl J. Friston, and Klaas E. Stephan. 2010. 'Observing the Observer (II): Deciding When to Decide'. *PloS One* 5(12):e15555. doi: 10.1371/journal.pone.0015555.

Daunizeau, Jean, Hanneke E. M. den Ouden, Matthias Pessiglione, Stefan J. Kiebel, Klaas E. Stephan, and Karl J. Friston. 2010. 'Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making'. *PLOS ONE* 5(12):e15554. doi: 10.1371/journal.pone.0015554.

Dębiec, Jacek, David E. A. Bush, and Joseph E. LeDoux. 2011. 'Noradrenergic Enhancement of Reconsolidation in the Amygdala Impairs Extinction of Conditioned Fear in Rats – a Possible Mechanism for the Persistence of Traumatic Memories in PTSD'. *Depression and Anxiety* 28(3):186–93. doi: 10.1002/da.20803.

Depue, B. E., T. Curran, and M. T. Banich. 2007. 'Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process'. *Science* 317(5835):215–19. doi: 10.1126/science.1139560.

Depue, B. E., J. M. Orr, H. R. Smolker, F. Naaz, and M. T. Banich. 2016. 'The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive, Emotional, and Motor Processes'. *Cerebral Cortex* 26(4):1634–46. doi: 10.1093/cercor/bhu324.

Desmedt, Aline, Aline Marighetto, and Pier-Vincenzo Piazza. 2015. 'Abnormal Fear Memory as a Model for Posttraumatic Stress Disorder'. *Biological Psychiatry* 78(5):290–97. doi: 10.1016/j.biopsych.2015.06.017.

Diaconescu, Andreea O., Christoph Mathys, Lilian A. E. Weber, Lars Kasper, Jan Mauer, and Klaas E. Stephan. 2017. 'Hierarchical Prediction Errors in Midbrain and Septum during Social Learning'. *Social Cognitive and Affective Neuroscience* 12(4):618–34. doi: 10.1093/scan/nsw171.

Duvernoy, Henri M. 2005. *The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections with MRI*. Springer Science & Business Media.

Ehlers, Anke. 2010. 'Understanding and Treating Unwanted Trauma Memories in Posttraumatic Stress Disorder'. *Zeitschrift Für Psychologie / Journal of Psychology* 218(2):141–45. doi: 10.1027/0044-3409/a000021.

Ehlers, Anke, and David M. Clark. 2000. 'A Cognitive Model of Posttraumatic Stress Disorder'. *Behaviour Research and Therapy* 38(4):319–45. doi: 10.1016/S0005-7967(99)00123-0.

Ehlers, Anke, Ann Hackmann, and Tanja Michael. 2004. 'Intrusive Re-experiencing in Post-traumatic Stress Disorder: Phenomenology, Theory, and Therapy'. *Memory* 12(4):403–15. doi: 10.1080/09658210444000025.

Engen, Haakon G., and Michael C. Anderson. 2018. 'Memory Control: A Fundamental Mechanism of Emotion Regulation'. *Trends in Cognitive Sciences* 22(11):982–95. doi: 10.1016/j.tics.2018.07.015.

Fawcett, Jonathan M., and Justin C. Hulbert. 2020. 'The Many Faces of Forgetting: Toward a Constructive View of Forgetting in Everyday Life'. *Journal of Applied Research in Memory and Cognition* 9(1):1–18. doi: 10.1016/j.jarmac.2019.11.002.

Feynman, Richard P. 1998. *Statistical Mechanics: A Set Of Lectures*. Avalon Publishing.

Frässle, Stefan, Ekaterina I. Lomakina, Lars Kasper, Zina M. Manjaly, Alex Leff, Klaas P. Pruessmann, Joachim M. Buhmann, and Klaas E. Stephan. 2018. 'A Generative Model of Whole-Brain Effective Connectivity'. *NeuroImage* 179:505–29. doi: 10.1016/j.neuroimage.2018.05.058.

Frässle, Stefan, Yu Yao, Dario Schöbi, Eduardo A. Aponte, Jakob Heinzle, and Klaas E. Stephan. 2018. 'Generative Models for Clinical Applications in Computational Psychiatry'. *WIREs Cognitive Science* 9(3):e1460. doi: 10.1002/wcs.1460.

Friston, K. J., L. Harrison, and W. Penny. 2003. 'Dynamic Causal Modelling'. *NeuroImage* 19(4):1273–1302. doi: 10.1016/S1053-8119(03)00202-7.

Friston, K. J., A. Mechelli, R. Turner, and C. J. Price. 2000. 'Nonlinear Responses in FMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics'. *NeuroImage* 12(4):466–77. doi: 10.1006/nimg.2000.0630.

Friston, Karl. 2009. 'The Free-Energy Principle: A Rough Guide to the Brain?' *Trends in Cognitive Sciences* 13(7):293–301. doi: 10.1016/j.tics.2009.04.005.

Friston, Karl. 2010. 'The Free-Energy Principle: A Unified Brain Theory?' *Nature Reviews Neuroscience* 11(2):127–38. doi: 10.1038/nrn2787.

Friston, Karl J. 2011. 'Functional and Effective Connectivity: A Review'. *Brain Connectivity* 1(1):13–36. doi: 10.1089/brain.2011.0008.

Friston, Karl J., Jean Daunizeau, James Kilner, and Stefan J. Kiebel. 2010. 'Action and Behavior: A Free-Energy Formulation'. *Biological Cybernetics* 102(3):227–60. doi: 10.1007/s00422-010-0364-z.

Friston, Karl J., A. David Redish, and Joshua A. Gordon. 2017. 'Computational Nosology and Precision Psychiatry'. *Computational Psychiatry* 1:2–23. doi: 10.1162/CPSY_a_00001.

Friston, Karl J., Klaas Enno Stephan, Read Montague, and Raymond J. Dolan. 2014. 'Computational Psychiatry: The Brain as a Phantastic Organ'. *The Lancet Psychiatry* 1(2):148–58. doi: 10.1016/S2215-0366(14)70275-5.

Gagne, Christopher, Peter Dayan, and Sonia J. Bishop. 2018. 'When Planning to Survive Goes Wrong: Predicting the Future and Replaying the Past in Anxiety and PTSD'. *Current Opinion in Behavioral Sciences* 24:89–95. doi: 10.1016/j.cobeha.2018.03.013.

Gagnepain, Pierre, Richard N. Henson, and Michael C. Anderson. 2014. 'Suppressing Unwanted Memories Reduces Their Unconscious Influence via Targeted Cortical Inhibition'. *Proceedings of the National Academy of Sciences* 111(13):E1310–19. doi: 10.1073/pnas.1311468111.

Gagnepain, Pierre, Justin Hulbert, and Michael C. Anderson. 2017. 'Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories'. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37(27):6423–41. doi: 10.1523/JNEUROSCI.2732-16.2017.

Gagnon, Stephanie A., and Anthony D. Wagner. 2016. 'Acute Stress and Episodic Memory Retrieval: Neurobiological Mechanisms and Behavioral Consequences: Acute Stress and Episodic Memory Retrieval'. *Annals of the New York Academy of Sciences* 1369(1):55–75. doi: 10.1111/nyas.12996.

Garfinkel, Sarah N., James L. Abelson, Anthony P. King, Rebecca K. Sripada, Xin Wang, Laura M. Gaines, and Israel Liberzon. 2014. 'Impaired Contextual Modulation of Memories in PTSD: An FMRI and Psychophysiological Study of Extinction Retention and Fear Renewal'. *Journal of Neuroscience* 34(40):13435–43. doi: 10.1523/JNEUROSCI.4287-13.2014.

Gravitz, Lauren. 2019a. 'The Forgotten Part of Memory'. *Nature* 571(7766):S12–14. doi: 10.1038/d41586-019-02211-5.

Gravitz, Lauren. 2019b. 'The Importance of Forgetting'. *Nature* 571:S12–14.

Guzowski, John F., James J. Knierim, and Edvard I. Moser. 2004. 'Ensemble Dynamics of Hippocampal Regions CA3 and CA1'. *Neuron* 44(4):581–84. doi: 10.1016/j.neuron.2004.11.003.

Hardt, Oliver, Karim Nader, and Lynn Nadel. 2013. ‘Decay Happens: The Role of Active Forgetting in Memory’. *Trends in Cognitive Sciences* 17(3):111–20. doi: 10.1016/j.tics.2013.01.001.

Harrington, Ralph. 2003. ‘On the Tracks of Trauma: Railway Spine Reconsidered’. *Social History of Medicine: The Journal of the Society for the Social History of Medicine* 16(2):209–23. doi: 10.1093/shm/16.2.209.

Harvey, A. G., and R. A. Bryant. 1998. ‘The Effect of Attempted Thought Suppression in Acute Stress Disorder’. *Behaviour Research and Therapy* 36(6):583–90. doi: 10.1016/s0005-7967(98)00052-7.

Hayes, Jasmeet Pannu, Kevin S. LaBar, Gregory McCarthy, Elizabeth Selgrade, Jessica Nasser, Florin Dolcos, VISN 6 Mid-Atlantic MIRECC workgroup, and Rajendra A. Morey. 2011. ‘Reduced Hippocampal and Amygdala Activity Predicts Memory Distortions for Trauma Reminders in Combat-Related PTSD’. *Journal of Psychiatric Research* 45(5):660–69. doi: 10.1016/j.jpsychires.2010.10.007.

Hebb, D. O. 1949. ‘The Organization of Behaviour’. *A Neurophysiological Theory*.

Heeger, David J., and David Ress. 2002. ‘What Does FMRI Tell Us about Neuronal Activity?’ *Nature Reviews Neuroscience* 3(2):142–51. doi: 10.1038/nrn730.

Heinzle, Jakob, and Klaas E. Stephan. 2018. ‘Chapter 5 - Dynamic Causal Modeling and Its Application to Psychiatric Disorders’. Pp. 117–44 in *Computational Psychiatry*, edited by A. Anticevic and J. D. Murray. Academic Press.

Hirst, William, and Elizabeth A. Phelps. 2016. ‘Flashbulb Memories’. *Current Directions in Psychological Science* 25(1):36–41. doi: 10.1177/0963721415622487.

Holmes, Emily A., Ata Ghaderi, Ellinor Eriksson, Klara Olofsdotter Lauri, Olivia M. Kukacka, Maya Mamish, Ella L. James, and Renée M. Visser. 2017. ‘“I Can’t Concentrate”: A Feasibility Study with Young Refugees in Sweden on Developing Science-Driven Interventions for Intrusive Memories Related to Trauma’. *Behavioural and Cognitive Psychotherapy* 45(2):97–109. doi: 10.1017/S135246581600062X.

Holmes, Emily A., Nick Grey, and Kerry A. D. Young. 2005. ‘Intrusive Images and “Hotspots” of Trauma Memories in Posttraumatic Stress Disorder: An Exploratory

Investigation of Emotions and Cognitive Themes'. *Journal of Behavior Therapy and Experimental Psychiatry* 36(1):3–17. doi: 10.1016/j.jbtep.2004.11.002.

Homan, Philipp, Ifat Levy, Eric Feltham, Charles Gordon, Jingchu Hu, Jian Li, Robert H. Pietrzak, Steven Southwick, John H. Krystal, Ilan Harpaz-Rotem, and Daniela Schiller. 2019. 'Neural Computations of Threat in the Aftermath of Combat Trauma'. *Nature Neuroscience* 22(3):470–76. doi: 10.1038/s41593-018-0315-x.

Hu, Sien, Jaime S. Ide, Sheng Zhang, and Chiang-shan R. Li. 2016. 'The Right Superior Frontal Gyrus and Individual Variation in Proactive Control of Impulsive Response'. *Journal of Neuroscience* 36(50):12688–96. doi: 10.1523/JNEUROSCI.1175-16.2016.

Hulbert, Justin C., Richard N. Henson, and Michael C. Anderson. 2016. 'Inducing Amnesia through Systemic Suppression'. *Nature Communications* 7(1):11003. doi: 10.1038/ncomms11003.

Hulbert, Justin C., Geeta Shivde, and Michael C. Anderson. 2012. 'Evidence against Associative Blocking as a Cause of Cue-Independent Retrieval-Induced Forgetting'. *Experimental Psychology* 59(1):11–21. doi: 10.1027/1618-3169/a000120.

Hunsaker, Michael R., and Raymond P. Kesner. 2013. 'The Operation of Pattern Separation and Pattern Completion Processes Associated with Different Attributes or Domains of Memory'. *Neuroscience and Biobehavioral Reviews* 37(1):36–58. doi: 10.1016/j.neubiorev.2012.09.014.

Hupbach, Almut, Rebecca Gomez, Oliver Hardt, and Lynn Nadel. 2007. 'Reconsolidation of Episodic Memories: A Subtle Reminder Triggers Integration of New Information'. *Learning & Memory* 14(1–2):47–53. doi: 10.1101/lm.365707.

Ide, Jaime S., Pradeep Shenoy, Angela J. Yu, and Chiang-shan R. Li. 2013. 'Bayesian Prediction and Evaluation in the Anterior Cingulate Cortex'. *Journal of Neuroscience* 33(5):2039–47. doi: 10.1523/JNEUROSCI.2201-12.2013.

Iglesias, Sandra, Christoph Mathys, Kay H. Brodersen, Lars Kasper, Marco Piccirelli, Hanneke E. M. den Ouden, and Klaas E. Stephan. 2013. 'Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning'. *Neuron* 80(2):519–30. doi: 10.1016/j.neuron.2013.09.009.

Jiang, Jiefeng, Katherine Heller, and Tobias Egner. 2014. 'Bayesian Modeling of Flexible Cognitive Control'. *Neuroscience and Biobehavioral Reviews* 46 Pt 1:30–43. doi: 10.1016/j.neubiorev.2014.06.001.

Kaouane, Nadia, Yves Porte, Monique Vallée, Laurent Brayda-Bruno, Nicole Mons, Ludovic Calandreau, Aline Marighetto, Pier Vincenzo Piazza, and Aline Desmedt. 2012. 'Glucocorticoids Can Induce PTSD-Like Memory Impairments in Mice'. *Science* 335(6075):1510–13. doi: 10.1126/science.1207615.

Karl, Anke, Michael Schaefer, Loretta S. Malta, Denise Dörfel, Nicolas Rohleder, and Annett Werner. 2006. 'A Meta-Analysis of Structural Brain Abnormalities in PTSD'. *Neuroscience & Biobehavioral Reviews* 30(7):1004–31. doi: 10.1016/j.neubiorev.2006.03.004.

van der Kolk, Bessel A. 2007. 'The History of Trauma in Psychiatry'. Pp. 19–36 in *Handbook of PTSD: Science and practice*. New York, NY, US: The Guilford Press.

Lawson, Rebecca P., Christoph Mathys, and Geraint Rees. 2017. 'Adults with Autism Overestimate the Volatility of the Sensory Environment'. *Nature Neuroscience* 20(9):1293–99. doi: 10.1038/nn.4615.

Layton, Barry, and Robert Krikorian. 2002. 'Memory Mechanisms in Posttraumatic Stress Disorder'. *The Journal of Neuropsychiatry and Clinical Neurosciences* 14(3):254–61. doi: 10.1176/jnp.14.3.254.

Levy, B. J., and M. C. Anderson. 2012a. 'Purging of Memories from Conscious Awareness Tracked in the Human Brain'. *Journal of Neuroscience* 32(47):16785–94. doi: 10.1523/JNEUROSCI.2640-12.2012.

Levy, B. J., and M. C. Anderson. 2012b. 'Purging of Memories from Conscious Awareness Tracked in the Human Brain'. *Journal of Neuroscience* 32(47):16785–94. doi: 10.1523/JNEUROSCI.2640-12.2012.

Levy, Benjamin J., and Anthony D. Wagner. 2011. 'Cognitive Control and Right Ventrolateral Prefrontal Cortex: Reflexive Reorienting, Motor Inhibition, and Action Updating'. *Annals of the New York Academy of Sciences* 1224(1):40–62. doi: 10.1111/j.1749-6632.2011.05958.x.

Lewandowsy, Stephan. 2010. ‘Forgetting in Memory Models: Arguments against Trace Decay and Consolidation Failure’. in *Forgetting*. Psychology Press.

Lindauer, Ramón J. L., Miranda Olff, Els P. M. van Meijel, Ingrid V. E. Carlier, and Berthold P. R. Gersons. 2006. ‘Cortisol, Learning, Memory, and Attention in Relation to Smaller Hippocampal Volume in Police Officers with Posttraumatic Stress Disorder’. *Biological Psychiatry* 59(2):171–77. doi: 10.1016/j.biopsych.2005.06.033.

Liu, Yunzhe, Wanjun Lin, Chao Liu, Yuejia Luo, Jianhui Wu, Peter J. Bayley, and Shaozheng Qin. 2016. ‘Memory Consolidation Reconfigures Neural Pathways Involved in the Suppression of Emotional Memories’. *Nature Communications* 7(1):13375. doi: 10.1038/ncomms13375.

Loughran, Tracey. 2012. ‘Shell Shock, Trauma, and the First World War: The Making of a Diagnosis and Its Histories’. *Journal of the History of Medicine and Allied Sciences* 67(1):94–119. doi: 10.1093/jhmas/jrq052.

Malenka, Robert C., and and Roger A. Nicoll. 1999. ‘Long-Term Potentiation--A Decade of Progress?’ *Science* 285(5435):1870–74. doi: 10.1126/science.285.5435.1870.

Maren, Stephen, K. Luan Phan, and Israel Liberzon. 2013. ‘The Contextual Brain: Implications for Fear Conditioning, Extinction and Psychopathology’. *Nature Reviews Neuroscience* 14(6):417–28. doi: 10.1038/nrn3492.

van Marle, Hein. 2015. ‘PTSD as a Memory Disorder’. *European Journal of Psychotraumatology* 6(1):27633. doi: 10.3402/ejpt.v6.27633.

Mary, Alison, Jacques Dayan, Giovanni Leone, Charlotte Postel, Florence Fraisse, Carine Malle, Thomas Vallée, Carine Klein-Peschanski, Fausto Viader, Vincent de la Sayette, Denis Peschanski, Francis Eustache, and Pierre Gagnepain. 2020. ‘Resilience after Trauma: The Role of Memory Suppression’. *Science* 367(6479). doi: 10.1126/science.aay8477.

Mathys, Christoph, Jean Daunizeau, Karl J. Friston, and Klaas Enno Stephan. 2011. ‘A Bayesian Foundation for Individual Learning Under Uncertainty’. *Frontiers in Human Neuroscience* 5. doi: 10.3389/fnhum.2011.00039.

McEwen, Bruce S., Carla Nasca, and Jason D. Gray. 2016. 'Stress Effects on Neuronal Structure: Hippocampus, Amygdala, and Prefrontal Cortex'. *Neuropsychopharmacology* 41(1):3–23. doi: 10.1038/npp.2015.171.

McGaugh, James L. 2000. 'Memory--a Century of Consolidation'. *Science* 287(5451):248–51. doi: 10.1126/science.287.5451.248.

Meeter, Martijn, and Jaap M. J. Murre. 2004. 'Consolidation of Long-Term Memory: Evidence and Alternatives'. *Psychological Bulletin* 130(6):843–57. doi: 10.1037/0033-2909.130.6.843.

Michael, T., A. Ehlers, S. L. Halligan, and D. M. Clark. 2005. 'Unwanted Memories of Assault: What Intrusion Characteristics Are Associated with PTSD?' *Behaviour Research and Therapy* 43(5):613–28. doi: 10.1016/j.brat.2004.04.006.

Mizumori, S. J., B. L. McNaughton, C. A. Barnes, and K. B. Fox. 1989. 'Preserved Spatial Coding in Hippocampal CA1 Pyramidal Cells during Reversible Suppression of CA3c Output: Evidence for Pattern Completion in Hippocampus'. *Journal of Neuroscience* 9(11):3915–28. doi: 10.1523/JNEUROSCI.09-11-03915.1989.

Mostofsky, Stewart H., and Daniel J. Simmonds. 2008. 'Response Inhibition and Response Selection: Two Sides of the Same Coin'. *Journal of Cognitive Neuroscience* 20(5):751–61. doi: 10.1162/jocn.2008.20500.

Moutoussis, Michael, Nitzan Shahar, Tobias U. Hauser, and Raymond J. Dolan. 2018. 'Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies'. *Computational Psychiatry* 2(0):50–73. doi: 10.1162/CPSY_a_00014.

Müller, Georg Elias, and Alfons Pilzecker. 1900. *Experimentelle Beiträge Zur Lehre Vom Gedächtniss*. Vol. 1. JA Barth.

Nietzsche, Friedrich. 1886. *Beyond Good and Evil*. Vintage.

North, Carol S., Alina M. Surís, Rebecca P. Smith, and Richard V. King. 2016. 'The Evolution of PTSD Criteria across Editions of DSM'. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists* 28(3):197–208.

Oitzl, Melly S., Marc Fluttert, and E. Ron de Kloet. 1998. 'Acute Blockade of Hippocampal Glucocorticoid Receptors Facilitates Spatial Learning in Rats'. *Brain Research* 797(1):159–62. doi: 10.1016/S0006-8993(98)00387-4.

Penny, Will D., Klaas E. Stephan, Jean Daunizeau, Maria J. Rosa, Karl J. Friston, Thomas M. Schofield, and Alex P. Leff. 2010. 'Comparing Families of Dynamic Causal Models' edited by K. P. Kording. *PLoS Computational Biology* 6(3):e1000709. doi: 10.1371/journal.pcbi.1000709.

Pezzulo, Giovanni, and Dimitri Ognibene. 2012. 'Proactive Action Preparation: Seeing Action Preparation as a Continuous and Proactive Process'. *Motor Control* 16(3):386–424. doi: 10.1123/mcj.16.3.386.

Phelps, Elizabeth A. 2004. 'Human Emotion and Memory: Interactions of the Amygdala and Hippocampal Complex'. *Current Opinion in Neurobiology* 14(2):198–202. doi: 10.1016/j.conb.2004.03.015.

Pitman, Roger K., Ann M. Rasmusson, Karestan C. Koenen, Lisa M. Shin, Scott P. Orr, Mark W. Gilbertson, Mohammed R. Milad, and Israel Liberzon. 2012. 'Biological Studies of Post-Traumatic Stress Disorder'. *Nature Reviews Neuroscience* 13(11):769–87. doi: 10.1038/nrn3339.

Pohlack, Sebastian T., Frauke Nees, Claudia Liebscher, Raffaele Cacciaglia, Slawomira J. Diener, Stephanie Ridder, Friedrich G. Woermann, and Herta Flor. 2011. 'Hippocampal but Not Amygdalar Volume Affects Contextual Fear Conditioning in Humans'. *Human Brain Mapping* 33(2):478–88. doi: 10.1002/hbm.21224.

Popoli, Maurizio, Zhen Yan, Bruce S. McEwen, and Gerard Sanacora. 2012. 'The Stressed Synapse: The Impact of Stress and Glucocorticoids on Glutamate Transmission'. *Nature Reviews Neuroscience* 13(1):22–37. doi: 10.1038/nrn3138.

Postel, Charlotte, Alison Mary, Jacques Dayan, Florence Fraisse, Thomas Vallée, Bérengère Guillery-Girard, Fausto Viader, Vincent de la Sayette, Denis Peschanski, Francis Eustache, and Pierre Gagnepain. 2021. 'Variations in Response to Trauma and Hippocampal Subfield Changes'. *Neurobiology of Stress* 15:100346. doi: 10.1016/j.ynstr.2021.100346.

Proust, Marcel. 1919. *À La Recherche Du Temps Perdu*. Vol. 1. Gallimard. Paris.

Qin, Shaozheng, Erno J. Hermans, Hein J. F. van Marle, Jing Luo, and Guillén Fernández. 2009. 'Acute Psychological Stress Reduces Working Memory-Related Activity in the Dorsolateral Prefrontal Cortex'. *Biological Psychiatry* 66(1):25–32. doi: 10.1016/j.biopsych.2009.03.006.

Rao, Rajesh P. N., and Dana H. Ballard. 1999. 'Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects'. *Nature Neuroscience* 2(1):79–87. doi: 10.1038/4580.

Reynolds, Martina, and Chris R. Brewin. 1999. 'Intrusive Memories in Depression and Posttraumatic Stress Disorder'. *Behaviour Research and Therapy* 37(3):201–15. doi: 10.1016/S0005-7967(98)00132-6.

Ribot, Théodule Armand. 1882. *Diseases of Memory: An Essay in the Positive Psychology*. New York: D. Appleton and company.

Rigoux, L., K. E. Stephan, K. J. Friston, and J. Daunizeau. 2014. 'Bayesian Model Selection for Group Studies — Revisited'. *NeuroImage* 84:971–85. doi: 10.1016/j.neuroimage.2013.08.065.

Rolls, Edmund. 2013. 'The Mechanisms for Pattern Completion and Pattern Separation in the Hippocampus'. *Frontiers in Systems Neuroscience* 7:74. doi: 10.3389/fnsys.2013.00074.

Rolls, Edmund T. 2010. 'A Computational Theory of Episodic Memory Formation in the Hippocampus'. *Behavioural Brain Research* 215(2):180–96. doi: 10.1016/j.bbr.2010.03.027.

Rudy, Jerry W., Ruth M. Barrientos, and Randall C. O'Reilly. 2002. 'Hippocampal Formation Supports Conditioning to Memory of a Context'. *Behavioral Neuroscience* 116(4):530–38. doi: 10.1037//0735-7044.116.4.530.

Rupprechter, Samuel, Aistis Stankevicius, Quentin J. M. Huys, J. Douglas Steele, and Peggy Seriès. 2018. 'Major Depression Impairs the Use of Reward Values for Decision-Making'. *Scientific Reports* 8(1):13798. doi: 10.1038/s41598-018-31730-w.

Schmitz, Taylor W., Marta M. Correia, Catarina S. Ferreira, Andrew P. Prescot, and Michael C. Anderson. 2017. 'Hippocampal GABA Enables Inhibitory Control over Unwanted Thoughts'. *Nature Communications* 8(1):1311. doi: 10.1038/s41467-017-00956-z.

Schwabe, Lars, Marian Joëls, Benno Roozendaal, Oliver T. Wolf, and Melly S. Oitzl. 2012. 'Stress Effects on Memory: An Update and Integration'. *Neuroscience & Biobehavioral Reviews* 36(7):1740–49. doi: 10.1016/j.neubiorev.2011.07.002.

Schwabe, Lars, Karim Nader, and Jens C. Pruessner. 2014. 'Reconsolidation of Human Memory: Brain Mechanisms and Clinical Relevance'. *Biological Psychiatry* 76(4):274–80. doi: 10.1016/j.biopsych.2014.03.008.

Semon, Richard. 1921. *The Mneme*. Allen and Unwin. London.

Seriès, Peggy. 2019. 'Post-Traumatic Stress Disorder as a Disorder of Prediction'. *Nature Neuroscience* 22(3):334–36. doi: 10.1038/s41593-019-0345-z.

Shields, Grant S., Matthew A. Sazma, and Andrew P. Yonelinas. 2016. 'The Effects of Acute Stress on Core Executive Functions: A Meta-Analysis and Comparison with Cortisol'. *Neuroscience & Biobehavioral Reviews* 68:651–68. doi: 10.1016/j.neubiorev.2016.06.038.

Shin, Lisa M., Scott L. Rauch, and Roger K. Pitman. 2006. 'Amygdala, Medial Prefrontal Cortex, and Hippocampal Function in PTSD'. *Annals of the New York Academy of Sciences* 1071:67–79. doi: 10.1196/annals.1364.007.

Siegel, Jenifer Z., Suzanne Estrada, Molly J. Crockett, and Arielle Baskin-Sommers. 2019. 'Exposure to Violence Affects the Development of Moral Impressions and Trust Behavior in Incarcerated Males'. *Nature Communications* 10(1):1942. doi: 10.1038/s41467-019-09962-9.

Sinclair, Alyssa H., and Morgan D. Barense. 2019. 'Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation'. *Trends in Neurosciences* 42(10):727–39. doi: 10.1016/j.tins.2019.08.007.

Smith, Michael E. 2005. 'Bilateral Hippocampal Volume Reduction in Adults with Post-Traumatic Stress Disorder: A Meta-Analysis of Structural MRI Studies'. *Hippocampus* 15(6):798–807. doi: 10.1002/hipo.20102.

Steiger, Frauke, Frauke Nees, Manon Wicking, Simone Lang, and Herta Flor. 2015. 'Behavioral and Central Correlates of Contextual Fear Learning and Contextual Modulation of Cued Fear in Posttraumatic Stress Disorder'. *International Journal of Psychophysiology* 98(3):584–93. doi: 10.1016/j.ijpsycho.2015.06.009.

Stein, Murray B., and Martin P. Paulus. 2009. 'Imbalance of Approach and Avoidance: The Yin and Yang of Anxiety Disorders'. *Biological Psychiatry* 66(12):1072–74. doi: 10.1016/j.biopsych.2009.09.023.

Stephan, K. E., W. D. Penny, R. J. Moran, H. E. M. den Ouden, J. Daunizeau, and K. J. Friston. 2010. 'Ten Simple Rules for Dynamic Causal Modeling'. *NeuroImage* 49(4):3099–3109. doi: 10.1016/j.neuroimage.2009.11.015.

Stephan, K. E., F. Schlagenhauf, Q. J. M. Huys, S. Raman, E. A. Aponte, K. H. Brodersen, L. Rigoux, R. J. Moran, J. Daunizeau, R. J. Dolan, K. J. Friston, and A. Heinz. 2017. 'Computational Neuroimaging Strategies for Single Patient Predictions'. *NeuroImage* 145:180–99. doi: 10.1016/j.neuroimage.2016.06.038.

Stephan, Klaas E., Sandra Iglesias, Jakob Heinzle, and Andreea O. Diaconescu. 2015. 'Translational Perspectives for Computational Neuroimaging'. *Neuron* 87(4):716–32. doi: 10.1016/j.neuron.2015.07.008.

Stephan, Klaas E., Zina M. Manjaly, Christoph D. Mathys, Lilian A. E. Weber, Saee Paliwal, Tim Gard, Marc Tittgemeyer, Stephen M. Fleming, Helene Haker, Anil K. Seth, and Frederike H. Petzschner. 2016. 'Allostatic Self-Efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression'. *Frontiers in Human Neuroscience* 10:550. doi: 10.3389/fnhum.2016.00550.

Stephan, Klaas Enno, Lars Kasper, Lee M. Harrison, Jean Daunizeau, Hanneke E. M. den Ouden, Michael Breakspear, and Karl J. Friston. 2008. 'Nonlinear Dynamic Causal Models for FMRI'. *NeuroImage* 42(2):649–62. doi: 10.1016/j.neuroimage.2008.04.262.

Stephan, Klaas Enno, and Christoph Mathys. 2014. 'Computational Approaches to Psychiatry'. *Current Opinion in Neurobiology* 25:85–92. doi: 10.1016/j.conb.2013.12.007.

Stephan, Klaas Enno, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. 2009. 'Bayesian Model Selection for Group Studies'. *NeuroImage* 46(4):1004–17. doi: 10.1016/j.neuroimage.2009.03.025.

Sullivan, Danielle R., Brian Marx, May S. Chen, Brendan E. Depue, Scott M. Hayes, and Jasmeet P. Hayes. 2019. 'Behavioral and Neural Correlates of Memory Suppression in PTSD'. *Journal of Psychiatric Research* 112:30–37. doi: 10.1016/j.jpsychires.2019.02.015.

Surget, A., and C. Belzung. 2021. 'Adult Hippocampal Neurogenesis Shapes Adaptation and Improves Stress Response: A Mechanistic and Integrative Perspective'. *Molecular Psychiatry*. doi: 10.1038/s41380-021-01136-8.

Thorndike, Edward L. 1913. *Educational Psychology*. New York: Teachers College, Columbia University.

TURING, A. M. 1950. 'I.—COMPUTING MACHINERY AND INTELLIGENCE'. *Mind* LIX(236):433–60. doi: 10.1093/mind/LIX.236.433.

Vossel, Simone, Christoph Mathys, Jean Daunizeau, Markus Bauer, Jon Driver, Karl J. Friston, and Klaas E. Stephan. 2014. 'Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed'. *Cerebral Cortex* 24(6):1436–50. doi: 10.1093/cercor/bhs418.

Waldhauser, Gerd T., Martin J. Dahl, Martina Ruf-Leuschner, Veronika Müller-Bamouh, Maggie Schauer, Nikolai Axmacher, Thomas Elbert, and Simon Hanslmayr. 2018. 'The Neural Dynamics of Deficient Memory Control in Heavily Traumatized Refugees'. *Scientific Reports* 8(1):13132. doi: 10.1038/s41598-018-31400-x.

Wang, Zhen, Thomas C. Neylan, Susanne G. Mueller, Maryann Lenoci, Diana Truran, Charles R. Marmar, Michael W. Weiner, and Norbert Schuff. 2010. 'Magnetic Resonance Imaging of Hippocampal Subfields in Posttraumatic Stress Disorder'. *Archives of General Psychiatry* 67(3):296–303. doi: 10.1001/archgenpsychiatry.2009.205.

Wegner, D. M. 1994. 'Ironic Processes of Mental Control'. *Psychological Review* 101(1):34–52. doi: 10.1037/0033-295x.101.1.34.

Wenzlaff, Richard M., and Daniel M. Wegner. 2000. ‘Thought Suppression’. *Annual Review of Psychology* 51(1):59–91. doi: 10.1146/annurev.psych.51.1.59.

Wiltgen, Brian J., Miou Zhou, Ying Cai, J. Balaji, Mikael Guzman Karlsson, Sherveen N. Parivash, Weidong Li, and Alcino J. Silva. 2010. ‘The Hippocampus Plays a Selective Role in the Retrieval of Detailed Contextual Memories’. *Current Biology* 20(15):1336–44. doi: 10.1016/j.cub.2010.06.068.

Wixted, John T. 2004. ‘The Psychology and Neuroscience of Forgetting’. *Annual Review of Psychology* 55(1):235–69. doi: 10.1146/annurev.psych.55.090902.141555.

Yassa, Michael A., and Craig E. L. Stark. 2011. ‘Pattern Separation in the Hippocampus’. *Trends in Neurosciences* 34(10):515–25. doi: 10.1016/j.tins.2011.06.006.

Yuen, Eunice Y., Jing Wei, Wenhua Liu, Ping Zhong, Xiangning Li, and Zhen Yan. 2012. ‘Repeated Stress Causes Cognitive Impairment by Suppressing Glutamate Receptor Expression and Function in Prefrontal Cortex’. *Neuron* 73(5):962–77. doi: 10.1016/j.neuron.2011.12.033.

Zelikowsky, Moriel, Stephanie Bissiere, and Michael S. Fanselow. 2012. ‘Contextual Fear Memories Formed in the Absence of the Dorsal Hippocampus Decay Across Time’. *Journal of Neuroscience* 32(10):3393–97. doi: 10.1523/JNEUROSCI.4339-11.2012.

Zlotnick, Caron, C. Laurel Franklin, and Mark Zimmerman. 2002. ‘Does “Subthreshold” Posttraumatic Stress Disorder Have Any Clinical Relevance?’ *Comprehensive Psychiatry* 43(6):413–19. doi: 10.1053/comp.2002.35900.

ANNEX

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Resilience after trauma: The role of memory suppression

Alison Mary, Jacques Dayan, Giovanni Leone, Charlotte Postel, Florence Fraisse, Carine Malle, Thomas Vallée, Carine Klein-Peschanski, Fausto Viader, Vincent de la Sayette, Denis Peschanski, Francis Eustache, Pierre Gagnepain*

INTRODUCTION: One of the fundamental questions in clinical neuroscience is why some individuals can cope with traumatic events, while others remain traumatized by a haunting past they cannot get rid of. The expression and persistence of vivid and distressing intrusive memories is a central feature of posttraumatic stress disorder (PTSD). Current understanding of PTSD links this persistence to a failure to reduce the fear associated with the trauma, a deficit rooted in the dysfunction of memory. In this study, we investigated whether this deficit may additionally be rooted in the disruption of the brain system that normally allows control over memory.

RATIONALE: To test this hypothesis in a laboratory setting, we implemented neutral and inoffensive intrusive memories paired with a reminder cue in a group of 102 individuals exposed to the 2015 Paris terrorist attacks and in a group of 73 nonexposed individuals (i.e., individuals who did not experience the attacks). The exposed group was composed of 55 individuals suffering from PTSD symptoms (denoted PTSD+) and 47 individuals showing no noticeable impairment after the trauma (denoted PTSD−). We used functional magnetic resonance imaging to measure how the dorsolateral prefrontal cortex (DLPFC), a core hub of the brain control system, regulated and suppressed memory activity during the reexperiencing of these intrusive memories. We focused our analyses on both the functional and causal dependency between control and memory neural circuits during attempts to suppress the reemergence of these intrusive memories.

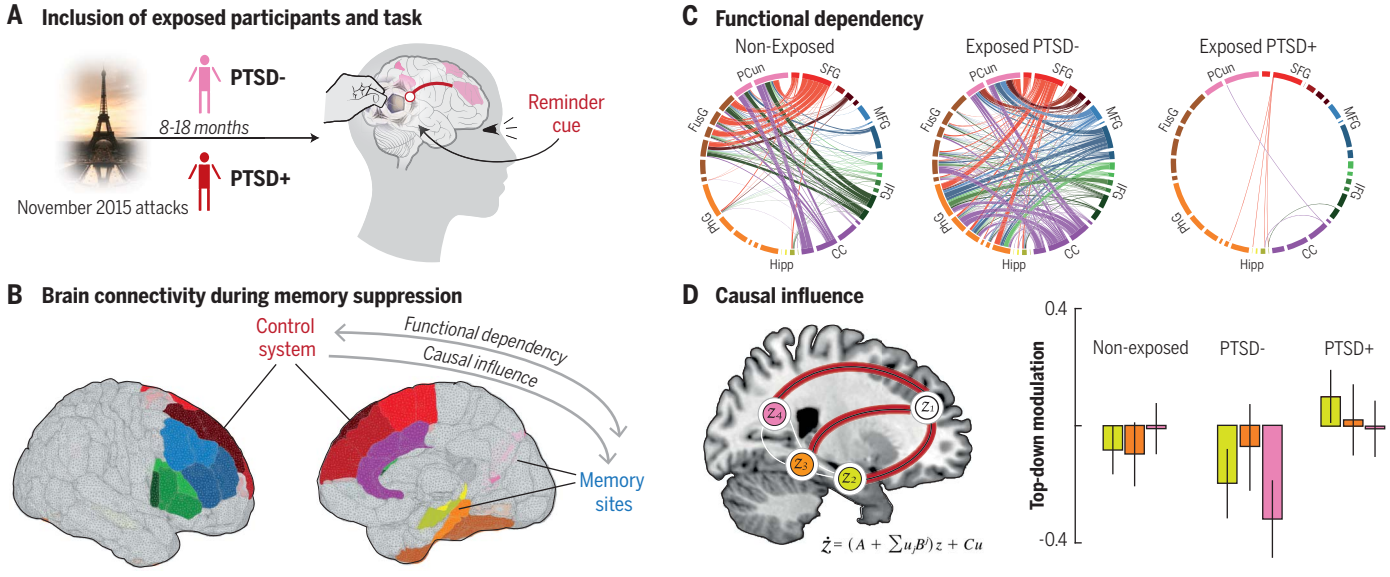
RESULTS: In healthy individuals (PTSD− and nonexposed), attempts to prevent the unwanted emergence of intrusive memory into consciousness was associated with a significant reduction of the functional coupling between control and memory systems, compared with situations where the reminder did not trigger such intrusion. In contrast, there was a near-absence of such a decrease in connectivity in PTSD+. Additional analyses focusing on the directionality of the underlying neural flow communications revealed that the suppression of intrusive memories in healthy individuals arose from the regulation of the right anterior DLPFC, which tuned the response of memory processes to reduce their responses. Notably, this regulation was directed at two key regions previously associated with the reexperiencing of traumatic memories: the hippocampus and the precuneus.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay8477

CONCLUSION: We observed a generalized disruption in PTSD of the regulation signal that controls the reactivation of unwanted memories. This disruption could constitute a central factor in the persistence of traumatic memories, undercutting the ability to deploy the necessary coping resources that maintain healthy memory. Such a deficit may explain maladaptive and unsuccessful suppression attempts often seen in PTSD. Our study suggests that the general mental operations typically engaged to banish and suppress the intrusive expression of unwanted memories might contribute to positive adaptation in the aftermath of a traumatic event, paving the way for new treatments. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: pierre.gagnepain@inserm.fr
Cite this article as A. Mary *et al.*, *Science* **367**, eaay8477 (2020). DOI: 10.1126/science.aay8477



Mechanisms of memory suppression after trauma. (**A**) Exposed individuals with or without PTSD were asked to suppress the reexperiencing of neutral intrusive memories. (**B**) Analyses focused on the functional and causal dependencies between control and memory systems during suppression attempts. (**C**) Extensive decreased coupling to counteract intrusion was seen in nonexposed and PTSD− groups but not in the PTSD+ group. SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; CC, cingulate cortex; Hipp, hippocampus; PhG, parahippocampal gyrus; FusG, fusiform gyrus; PCun, precuneus. (**D**) This decreased coupling was mediated by top-down regulation of involuntary memory processing arising from the right DLPFC.

RESEARCH ARTICLE

NEUROSCIENCE

Resilience after trauma: The role of memory suppression

Alison Mary[1], Jacques Dayan[1,2], Giovanni Leone[1], Charlotte Postel[1], Florence Fraisse[1], Carine Malle[1], Thomas Vallée[1], Carine Klein-Peschanski[3], Fausto Viader[1], Vincent de la Sayette[1], Denis Peschanski[3], Francis Eustache[1], Pierre Gagnepain[1]*

In the aftermath of trauma, little is known about why the unwanted and unbidden recollection of traumatic memories persists in some individuals but not others. We implemented neutral and inoffensive intrusive memories in the laboratory in a group of 102 individuals exposed to the 2015 Paris terrorist attacks and 73 nonexposed individuals, who were not in Paris during the attacks. While reexperiencing these intrusive memories, nonexposed individuals and exposed individuals without posttraumatic stress disorder (PTSD) could adaptively suppress memory activity, but exposed individuals with PTSD could not. These findings suggest that the capacity to suppress memory is central to positive posttraumatic adaptation. A generalized disruption of the memory control system could explain the maladaptive and unsuccessful suppression attempts often seen in PTSD, and this disruption should be targeted by specific treatments.

The expression and persistence of vivid, uncontrollable, and distressing intrusive memories is a central feature of posttraumatic stress disorder (PTSD) (*1*–*5*). After a traumatic event, attempts to suppress or avoid traumatic memories sometimes paradoxically increase the expression of intrusive memories (*6*–*8*). Successful treatments of intrusive memories involve overcoming such avoidance and suppression, as well as bringing back elements of the traumatic memory to promote its extinction or updating by the integration of a safe context (*2*, *5*, *9*, *10*). These treatments are in line with current neurobiological models that link PTSD to a learning impairment together with a deficit in processing contextual reminders in the fear circuit (*11*–*15*).

Theories of PTSD implicate experiential avoidance of traumatic memories via thought suppression as detrimental and central to the maintenance of intrusion symptoms (*2*, *16*–*19*). Experiential avoidance is mediated by the tonic maintenance of the to-be-avoided mental image in mind and by the engagement of a reactive inhibitory control process suppressing the momentary awareness of that unwanted thought (*20*, *21*). The former explains the paradoxical and maladaptive persistence of suppressed thoughts in memory and is exacerbated in PTSD (*22*, *23*). The latter, however, ultimately leads to forgetting of the suppressed event in healthy individuals (*24*–*31*).

[1]Normandie Université, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, GIP Cyceron, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France. [2]Pôle Hospitalo-Universitaire de Psychiatrie de l'Enfant et de l'Adolescent, Centre Hospitalier Guillaume Régnier, Université Rennes 1, 35700 Rennes, France. [3]Université Paris I Panthéon Sorbonne, HESAM Université, EHESS, CNRS, UMR8209, 75231 Paris, France.
*Corresponding author. Email: pierre.gagnepain@inserm.fr

Asking people to suppress awareness of a memory triggered by a reminder cue, without appealing to that memory, can impair its later conscious recall (*30*, *31*), unconscious expression (*27*, *32*, *33*), or emotional response (*34*, *35*). Memory suppression engages control mechanisms implemented by the frontoparietal network (*25*–*30*). Suppressing memory retrieval reduces activity over an extended network (*25*–*29*, *34*, *36*–*38*). Neurobiological models of motivated forgetting (*31*, *39*–*41*) assume that inhibitory control of memory awareness adaptively suppresses memory processing once retrieval cues have triggered interfering activity associated with unexpected intrusions. Suppression of hippocampal activity increases when unwanted memories intrude into awareness and need to be purged reactively (*34*, *36*, *37*). The central mechanisms associated with memory suppression are manifested as a negative influence of the right dorsolateral prefrontal cortex (DLPFC), especially the anterior middle frontal gyrus (MFG), over brain areas supporting the reactivation of memories (*26*, *27*). Such top-down suppression increases to adaptively counteract and regulate intrusion involuntarily emerging into a person's awareness (*34*, *36*).

Alteration of these inhibitory control mechanisms could represent a potentially critical mechanism underlying intrusive symptoms in PTSD that contributes to adverse outcomes. Thus, the perseveration of intrusive memories in PTSD after suppression attempts may arise from the existence of a compromised and ineffective memory control system. Disruption of the system controlling memories undercuts the ability to deploy the otherwise necessary coping skill of suppression. Any attempt to regulate and suppress intrusive memories is therefore doomed to failure and reflects futile efforts to slam on a faulty brake. This hypothesis receives support from behavioral and neural evidence for inhibitory control deficits in PTSD (*42*–*47*).

In this study, we measured the connectivity between the control system and memory circuits using functional magnetic resonance imaging (fMRI) in 102 exposed and 73 nonexposed individuals of the 13 November 2015 Paris terrorist attacks (see materials and methods for type of traumatic exposure, "nonexposed" meaning not present in Paris), while they attempted to suppress neutral and inoffensive intrusive memories implemented in the laboratory (Fig. 1B). Trauma-exposed participants (see table S1 for demographic and clinical characteristics) were divided into two groups: one group with full or partial symptomology of PTSD (*48*) according to current *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*) criteria (n = 55 individuals), and one group without PTSD (n = 47 individuals; see Fig. 1A and the materials and methods section). After learning word-object pairs, participants tried to stop the memory of the object from entering their awareness ("no-think") during the think/no-think (TNT) phase (Fig. 1B), which also included trials for which they had to recall the associated object ("think"). If the object came to mind anyway during suppression attempts, they were asked to push it out of mind and to report after the end of the trial that the reminder elicited awareness of its paired object (*37*), allowing us to isolate when no-think trials triggered intrusions.

Behavioral performances

In healthy individuals, intrusion decreases with repeated suppression of unwanted memory retrieval (*34*, *36*, *37*). Participants' control over intrusions improved across suppression repetitions in all three groups (Fig. 1C). A group times repetition analysis of variance (ANOVA) on participants' intrusion reports for no-think trials revealed a robust reduction in intrusion proportion with repetition [$F_{7,1204}$ = 30.3, P < 0.001]. Repeated suppressions reduced intrusions comparably for all three groups (group times repetition interaction was not significant) [$F_{14,1204}$ = 0.46, P = 0.95], and the overall proportion of intrusion did not differ between groups [$F_{2,172}$ = 2.1, P = 0.125].

After the TNT phase, we tested how easily participants could identify the objects amid visual noise. The amount of priming was reduced for no-think objects that were identified more slowly than objects from the baseline condition in nonexposed [t_{72} = 1.96, P = 0.027] and exposed non-PTSD [t_{46} = 1.73, P = 0.045] participants (see table S2 for mean reaction times and standard deviations). When objects reappeared in their visual world, participants found it harder to perceive suppressed objects than other recently encountered objects. This

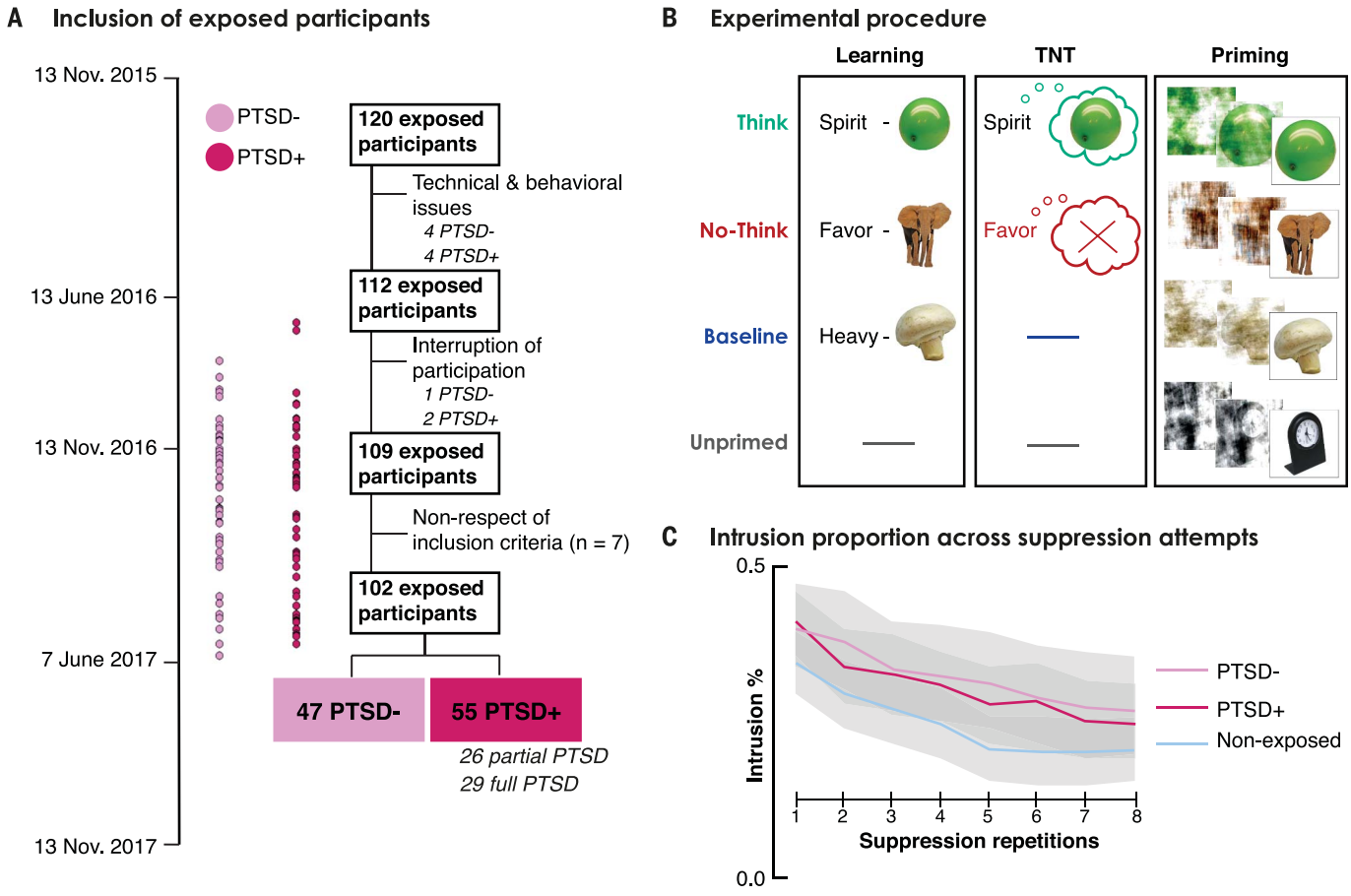


Fig. 1. Experimental design. (**A**) Timeline and procedure of inclusion of the participants exposed to the 13 November 2015 Paris terrorist attacks. The dates of the first and last inclusion are 13 June 2016 and 7 June 2017, respectively. Participants with a similar degree of exposure were diagnosed as non-PTSD or PTSD. (**B**) After learning word-object pairs, participants underwent fMRI scanning as they performed the think/no-think (TNT) task. For think items (in green), participants recalled a detailed visual memory of the associated picture. For no-think items (in red), they were asked to prevent the picture from entering awareness. After no-think trial cues ended, participants reported the presence or absence of intrusive memories that further trigger reactive inhibitory process. At the behavioral level, the effect of suppression was measured using a perceptual identification task including novel unprimed objects. (**C**) Intrusion proportions (i.e., the proportion of trials in which the associated memory entered into awareness on no-think trials) as measured by our trial-by-trial intrusion report measure (see materials and methods) over the eight suppression attempts of the TNT phase. Shaded error bands represent 95% bootstrapped confidence intervals.

reduction of priming effect after memory suppression was not found in the PTSD group [t_{54} = −0.84, P = 0.4], and the magnitude of this effect was significantly larger for the non-PTSD [t_{100} = 1.85, P = 0.033] and nonexposed [t_{126} = 1.95, P = 0.027] groups compared with the PTSD group, as shown by two-sample t tests. This difference could not be explained by a difference in training. Our procedure carefully matched learning of word-object associations, and no group differences emerged in the final criterion test before TNT procedure (correct recall: nonexposed, 93%; non-PTSD, 90%; and PTSD, 92%). Suppression-induced forgetting of explicit memories is impaired in PTSD (*44*). Our findings extend this deficit to perceptual implicit memory.

Brain activity

We first contrasted whole-brain activity of no-think and think trials. For all three groups, we observed the engagement of the right fronto-parietal control network (FPCN) and the disengagement of visual and medial temporal lobe (MTL) areas during retrieval suppression (fig. S1 and table S3). No noticeable differences were seen between non-PTSD and PTSD groups. We observed, however, a significant interaction when the trauma-exposed group with PTSD was compared to the nonexposed group. This interaction was observed using family-wise error (FWE) rate correction when the search volume was restricted to the FPCN (no-think greater than think contrast) and was driven by a greater engagement of the right superior frontal gyrus in the nonexposed group [Montreal Neurological Institute (MNI) coordinates: x = 16, y = 36, z = 56; Z = 4.34, $P_{\text{FWE-FPCN}}$ = 0.002]. It is unclear whether the ability to modulate and engage this region is disrupted by the existence of PTSD, or by trauma exposure rather than PTSD (*49*). This interaction might also reflect the daily engagement of trauma-exposed individuals in memory control processes and some form of habituation. Cortical thickness increases in a similar region after exposure to trauma, an effect that could potentially be related to experience-induced plasticity and habituation (*50*).

We next sought to examine whether people's ability to suppress intrusive memories depends on the engagement of the FPCN (*34*). The overall proportion of intrusions was entered into a regression model predicting the up-regulation of the control network during intrusion versus nonintrusion. The up-regulation of the fronto-parietal network was associated with a reduced intrusion frequency in both the nonexposed and non-PTSD groups (fig. S2). This relationship, however, was not observed in the exposed group of participants with PTSD.

Previous studies have observed more pronounced down-regulation of hippocampal

activity during retrieval suppression when memories involuntarily intrude into consciousness compared with when they do not (*34*, *36*, *37*). Although we observed a suppression-induced reduction of bilateral hippocampal activity in all three groups (nonexposed: [t_{72} = 4.78, P < 0.001]; non-PTSD: [t_{46} = 6.8, P < 0.001]; PTSD: [t_{54} = 5.67, P < 0.001]), no additional modulation was caused by the elevated control demand associated with intrusions (all P > 0.25) (fig. S3A). We did find more pronounced suppression of hippocampal activity in response to intrusion in all three groups (fig. S3B), but only when an adaptive volume restricted to the most significant contiguous voxels associated with the main effect of suppression was used (*34*). Outside the hippocampus, the suppression of intrusion in the two exposed groups, but not in the nonexposed group, was associated with a decrease over the lateral and posterior regions of the visual system (tables S5 to S7). However, no interaction between groups was observed. No noticeable differences in suppression strategy were observed between groups (fig. S4) (see materials and methods).

Functional connectivity

Next, we investigated the pattern of functional connectivity between the inhibitory control network and memory areas for the three groups (see materials and methods) (Fig. 2A and table S8). For the control network, we focused on the right-lateralized DLPFC (*25*–*30*), as well as the anterior cingulate cortex for its presumed role of relay in the DLPFC-hippocampal pathway (*41*). For the memory network, we included bilateral regions known to be modulated by inhibitory control mechanism and reflecting different memory domains (*25*–*30*, *34*, *36*, *37*).

We used a general linear regression model (GLM) and generalized psychophysiological interaction (gPPI) (*51*) to estimate task-dependent functional connectivity (between each pair of control-memory regions) across this broad network, while controlling for task-based activation and task-independent (i.e., physiological) functional connectivity. PPI was conducted with the inhibitory control network as seeds (i.e., independent variable of the regression model) and memory-related sites as target regions (i.e., dependent variable). We first characterized TNT-dependent functional connectivity changes for each group separately, focusing on significant changes between intrusion and nonintrusion. Inhibitory control models predict that intrusions will generate more negative coupling between frontally mediated control processes and memory regions (*31*, *40*, *41*). In the context of the current PPI analysis, this process would manifest as decreased connectivity during intrusion relative to nonintrusion. For both nonexposed and exposed non-PTSD groups, attempts to prevent the unwanted emergence of intrusive memory into consciousness were associated with a significant reduction in functional connectivity compared with nonintrusion in a broad network (Fig. 2B). These changes were characterized by a decrease in connectivity during intrusion (compared with nonintrusion) between an extensive frontal network and the parahippocampal gyrus, hippocampus, fusiform gyrus, and precuneus. When memories intruded awareness and needed to be purged, there was a near-absence of such a decrease in the connectivity in the exposed PTSD group (Fig. 2B).

However, these analyses did not formally establish that healthy and PTSD participants rely on different processes to suppress memory, which requires demonstrating the presence of a significant pattern of interaction between memory awareness (i.e., intrusive versus nonintrusive memories) and the groups. We thus focused on the connectivity changes between the right anterior MFG and memory regions (see materials and methods and Fig. 2A). The right anterior MFG region is critical for inhibitory control in a variety of cognitive task contexts (*28*) and inhibitory regulation of conscious awareness for unwanted memories (*25*–*30*, *34*, *36*). After computing the difference in connectivity between intrusion and nonintrusion, we looked at the connectivity separately for each target region and hemisphere to identify which memory processing was preferentially targeted by inhibitory control, controlling for the expected proportion of type I error across multiple regions of interest (ROIs) using the false discovery rate (FDR) correction. Two-sample t tests showed that the reduction in connectivity for intrusion compared with nonintrusion was significantly greater for exposed participants without PTSD than for the PTSD group in the right rostral hippocampus [t_{100} = −1.9, P_{FDR} = 0.043]; the left [t_{100} = −4.09, P_{FDR} = 0.0004] and right [t_{100} = −2.24, P_{FDR} = 0.023] parahippocampal gyrus; the left [t_{100} = −2.3, P_{FDR} = 0.02] and right [t_{100} = −3.27, P_{FDR} = 0.004] fusiform gyrus; and the left [t_{100} = −2.71, P_{FDR} = 0.011] and right [t_{100} = −2.69, P_{FDR} = 0.011] precuneus. These differences were driven by significant decreases in connectivity for intrusive relative to nonintrusive memories in the non-PTSD group, as revealed by one-sample t tests (Fig. 3 and tables S9 and S10). These decreases were absent in the PTSD group (all P_{FDR} > 0.2) or reversed with an up-regulation in the left parahippocampal gyrus [t_{54} = 2.91, P = 0.026] and the right fusiform gyrus [t_{54} = 2.44, P = 0.045]. These latter effects in the PTSD group became marginal after FDR corrections (P_{FDR} = 0.053 and 0.09, respectively). The differences in connectivity seen for the non-PTSD group compared with the PTSD group were independent of type or duration of traumatic exposure, age, sex, education, or medication (table S11).

The pattern of results was less clear-cut for the nonexposed control group. We observed significant reduction in connectivity during intrusions compared with nonintrusion in the left [t_{72} = −2.37, P = 0.01] and right [t_{72} = −2.64, P = 0.005] precuneus that became a trend after FDR correction for multiple comparisons (P_{FDR} = 0.051). We also observed in the nonexposed control group a trend in the right rostral hippocampus [t_{72} = −1.496, P = 0.07] that did not survive FDR correction for multiple comparisons. When compared with the PTSD group, nonexposed control participants had a significantly greater reduction in connectivity for intrusion versus nonintrusion in the left parahippocampal gyrus [t_{126} = −1.76, P = 0.04]; the left [t_{126} = −1.76, P = 0.04] and right [t_{126} = −2.07, P = 0.02] fusiform gyrus; the left [t_{126} = −2.71, P = 0.003] and right [t_{126} = −2.31, P = 0.01] precuneus; and showed a trend in the right rostral hippocampus [t_{126} = −1.5, P = 0.068]. After FDR corrections, only the difference for the left precuneus was significant (P_{FDR} = 0.038), the difference for the right rostral hippocampus did not survive to correction (P_{FDR} = 0.1), and the differences in the other regions became marginal (P_{FDR} > 0.056) (table S10). After an additional analysis controlling for age, sex, education, and medication, using FDR correction for multiple comparisons, the difference between the nonexposed and PTSD groups remained significant in the left precuneus (table S11). It is often observed that a healthy population is composed of a mixture of people with good and poor control abilities, as reflected in distinct connectivity profiles (*27*, *34*, *36*). Furthermore, it is possible that nonexposed individuals continuously engaged the anterior MFG to suppress memory activity regardless of whether an intrusion was present.

Active versus resting-state connectivity

Inhibitory control models predict that memory suppression will generate more negative coupling between frontally mediated control processes and memory regions. Although this would manifest as decreased connectivity during intrusion relative to nonintrusion in PPI analysis, our design does not allow us to estimate absolute change in connectivity for isolated conditions (see materials and methods).

We therefore compared isolated indexes of task-dependent connectivity for each condition to a resting-state session collected after the TNT task. This approach relied on blind deconvolution to detect spontaneous event-related changes in the resting-state signal (*52*). From these pseudo-events, a gPPI regression model was recreated with parameter estimates

quantitatively comparable to TNT-dependent connectivity estimates (see materials and methods). Using these estimates of resting-state connectivity as a baseline, we found an active reduction in coupling between an extended right DLPFC network and memory areas in reaction to intrusions for both nonexposed and non-PTSD groups (fig. S5). The PTSD group exhibited a similar decrease in the DLPFC-to–memory system connectivity but mostly during nonintrusion trials. Notably, the nonexposed group also exhibited a reduction in connectivity during nonintrusion trials, in line with the idea that this group suppressed memory activity regardless of the presence or absence

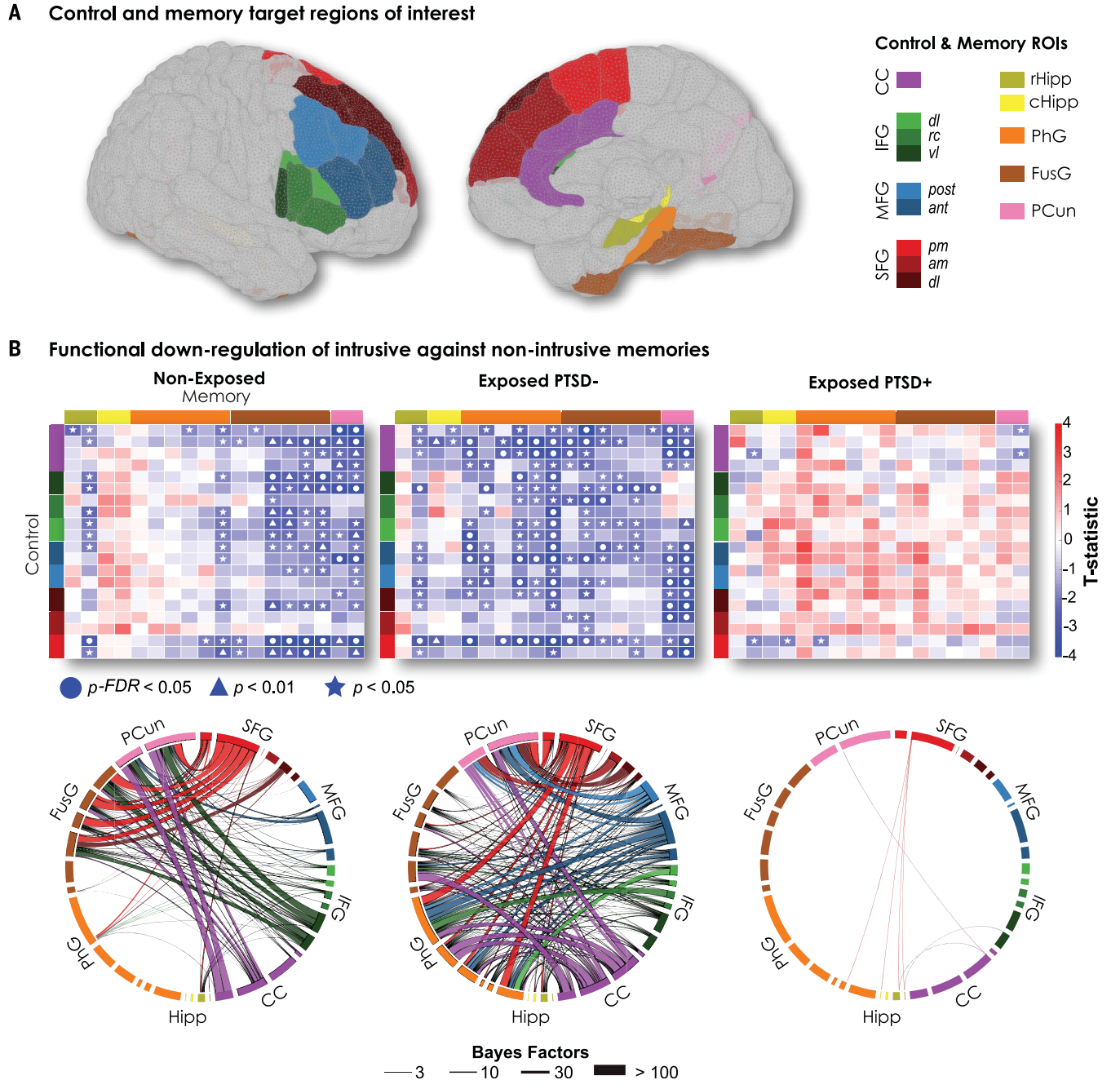


Fig. 2. Decrease in functional connectivity during suppression of intrusive memories between control and memory brain regions. (**A**) Suppression-induced functional connectivity was analyzed between prefrontal control (seed) and memory (target) regions of interest (ROIs). The control and memory target ROIs are represented as shown in the color key on the right. (**B**) The contrast between intrusion and nonintrusion shows an extensive decrease in connectivity for both the nonexposed and non-PTSD groups. The matrices represent connectivity changes (*t*-statistic) in each group, between the ROIs of the control and memory systems. Circles, triangles, and stars in the matrices represent significant changes in connection at $P_{FDR} < 0.05$, $P < 0.01$, and $P < 0.05$, respectively. In the circular connectograms, the colors of the edges are defined by the prefrontal control ROIs that predicted activity of memory sites in the gPPI model [color key in (A) applies here]. The size of the edges reflects the Bayes factors for connections associated with a significant decrease in connectivity during the regulation of intrusive compared with nonintrusive memories. SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; CC, cingulate cortex; Hipp, hippocampus; rHipp, rostral hippocampus; cHipp, caudal hippocampus; PhG, parahippocampal gyrus; FusG, fusiform gyrus; PCun, precuneus; pm, posterior medial; am, anterior medial; post, posterior; ant, anterior; dl, dorsolateral; rc, rostrocaudal; vl, ventrolateral.

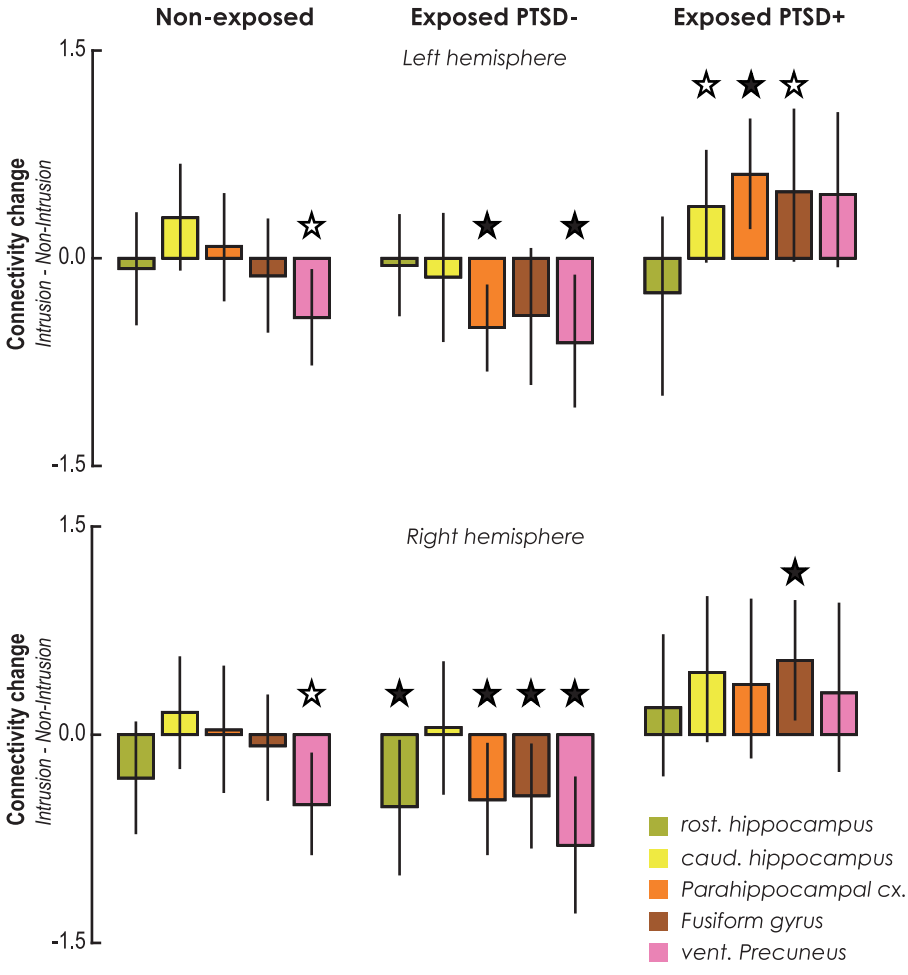


Fig. 3. Connectivity modulation between right anterior MFG and memory systems during memory suppression. Connectivity differences during the suppression of intrusive versus nonintrusive memories, between the right anterior MFG (seed) and target memory regions in the left (top panel) and right (bottom panel) hemispheres. Error bars reflect 95% bootstrapped confidence intervals and indicate significance when they do not encompass zero. Black and white stars indicate $P_{FDR} < 0.05$ and $P < 0.05$, respectively. rost., rostral; caud., caudal; cx., cortex; vent., ventral.

of intrusion. Focusing this analysis on the right anterior MFG revealed that the connectivity with memory sites, including the hippocampus, was reduced actively during intrusion in both non-PTSD and nonexposed groups (Fig. 4; see tables S12 and S13 for details on statistics). Such active reduction in connectivity was also found during nonintrusion trials in the left and right fusiform gyrus and right caudal hippocampus for the nonexposed group, as well as in the left parahippocampal gyrus and right fusiform gyrus for the exposed PTSD group (although these effects did not survive correction for multiple comparisons across tested memory areas). In the non-PTSD group, the decreased connectivity induced by memory suppression between control and memory systems reflected an active process that increased when intrusive memories arose into consciousness and needed to be purged. Also, no active differences in connectivity were found when reminder cues did not trigger intrusion in this group. These findings fit well with current neurobiological models of motivated forgetting (*39*–*41*), which propose that inhibitory control of memory adaptively increases to suppress memory processing once retrieval cues unexpectedly trigger interfering intrusive activity.

Top-down versus bottom-up connectivity

We used dynamic causal modeling (DCM) to analyze top-down and bottom-up influences separately during attempts to down-regulate intrusive memory. Because DCM is limited to a restricted number of nodes, we designed simple four-node DCM models to study the change in connectivity between the right anterior MFG on one hand, and the right rostral hippocampus, parahippocampal cortex, and precuneus on the other hand. We estimated seven models, reflecting possible differences in coupling between intrusion and nonintrusion trials (Fig. 5A), as well as an additional model without modulation (see materials and methods).

All three groups showed strong evidence for the presence of suppression-induced modulation of the connectivity between the right MFG and memory systems (see materials and methods). We used Bayesian model averaging (BMA) to weight the change in coupling parameters according to posterior model evidence across all seven possible combinations of modulation between MFG and memory targets (Fig. 5B). Down-regulation of intrusive memory activity in the rostral hippocampus was mediated by a top-down modulation (M) of the right anterior MFG in non-PTSD participants [M = −0.198; posterior probability (PP) = 0.997; 95% confidence interval (CI) = [−0.32, −0.08]] and nonexposed participants (M = −0.083; PP = 0.95; 95% CI = [−0.16, −0.0001]). Critically, such top-down modulation of involuntary memory processing in the rostral hippocampus was absent in the PTSD group, which exhibited the reversed pattern characterized by a greater decrease in MFG-to-hippocampus coupling during nonintrusion (M = 0.10; PP = 0.965; 95% CI = [0.009, 0.19]). Significant group-differences (Δ) between the PTSD group and both the non-PTSD (Δ = −0.30; PP = 0.999; 95% CI = [−0.45, −0.15]) and nonexposed (Δ = −0.18; PP = 0.95; 95% CI = [−0.31, −0.06]) groups were seen on top-down coupling parameters between the right MFG and rostral hippocampus. The non-PTSD group also showed a strong down-regulation of the precuneus (M = −0.30; PP = 0.999; 95% CI = [−0.45, −0.15]), an effect that was much stronger than the one seen in both PTSD (Δ = −0.31; PP = 0.999; 95% CI = [−0.49, −0.15]) and nonexposed (Δ = −0.32; PP = 1.0; 95% CI = [−0.48, −0.16]) groups. The differences in top-down connectivity seen for the non-PTSD group compared with the other two groups was independent of type or duration of traumatic exposure, age, sex, education, or medication (table S14).

A general deficit in the inhibitory control of intrusive memories in PTSD

Current models of PTSD link the persistence of intrusive memories to a failure of the extinction or updating of the original traumatic memory traces while in a safe environment, together with an abnormal and exaggerated processing of contextual reminder of the trauma in the fear circuit (*11*–*15*). These disruptions involve the dysfunction of the hippocampus-amygdala complex and its interaction with the medial prefrontal cortex. Our findings suggest that PTSD is also characterized by a deficit in the top-down suppression of momentary awareness associated with intrusive memories. This deficit could constitute a central factor in the persistence of traumatic memories, undercutting the ability to deploy

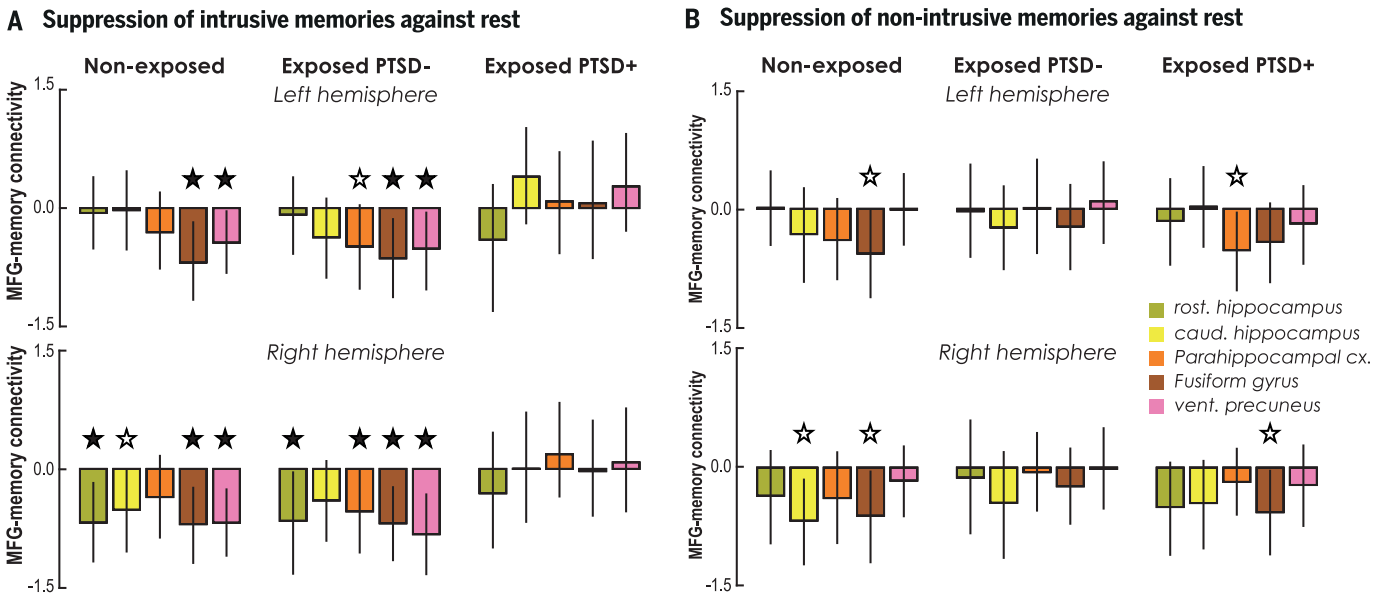


Fig. 4. Suppression-induced connectivity against rest. Connectivity differences induced by the suppression of intrusive (**A**) and nonintrusive (**B**) memories against a resting-state baseline, using the right anterior MFG as seed and memory regions as targets. Error bars reflect 95% bootstrapped confidence intervals and indicate significance when they do not encompass zero. Black and white stars indicate $P_{FDR} < 0.05$ and $P < 0.05$, respectively.

the necessary coping resources that maintain a healthy memory.

In trauma-exposed individuals without PTSD, the functional connectivity between prefrontal areas involved in control and memory sites, including the hippocampus and precuneus, decreased during the regulation of intrusive memory compared with nonintrusion. This decrease in connectivity was also seen in comparison to a resting-state baseline, suggesting that changes in connectivity induced by the suppression of intrusion relied on an active modulation. Analysis of effective connectivity showed that a top-down process mediated these modulations in non-PTSD, and that this effect was accentuated compared with PTSD. The current findings are consistent with the existence of an inhibitory signal that interrupts the reactivation of unwanted memory traces in memory systems (*29*, *34*). Such inhibitory control was preserved in resilient individuals but disrupted in people who developed PTSD.

The intrusive memories created in the current experiment are completely different from the distressing, fragmented, and decontextualized traumatic intrusions seen in PTSD (*1*–*5*). However, common features that are central to PTSD symptomatology also exist and can be modeled and isolated using the TNT paradigm. Both types of intrusions are involuntary, unintended, composed of sensory impressions, and triggered by unrelated contextual cues weakly related to the memory content (*2*). Neutral memories completely unrelated to the traumatic event also put exposed and nonexposed individuals on equal footing regarding the control demand associated with memory intrusion. Moreover, the regulation of neutral and emotional memories is probably achieved by the same core control system (*25*, *28*, *34*). Our findings thus highlight the presence of a central and general disruption of the down-regulation function of the anterior DLPFC in PTSD, disrupting the control and suppression of involuntarily intruding memories, even when those memories are neutral, artificially created, equated in strength during learning, and completely unrelated to the traumatic event.

Suppressing memories is often assumed to be unwise because the undesired remnants will backfire (*2*, *6*–*8*, *16*–*19*). Rather than being the root of intrusive symptoms, our findings suggest that maladaptive and unsuccessful suppression attempts are a consequence of a compromised control system. Such disruption may prevent adaptive forgetting processes (*31*) that normally alter memory stabilization in the hippocampus (*38*) and might therefore prevent the impairment of the traumatic engram. Furthermore, alteration of control capacity can further cascade into an exaggerated avoidance of reminders of the trauma. Unlike memory suppression, avoidance of reminders prevents modulation of traumatic representations via inhibitory control (*53*), extinction, or updating (*13*–*15*). Disrupted inhibitory control processes could accentuate the imbalance between memory suppression and avoidance strategies, which reflect the same goal of keeping the trauma memory out of awareness but have opposite consequences on mental health.

Inhibitory control: Resilience or vulnerability to PTSD?

Do such inhibitory control mechanisms engaged during memory suppression reflect a preexisting resilience factor, some form of positive and dynamic adaptation after exposure to a traumatic event, a preexisting vulnerability factor, or sequelae exacerbated by chronic stress (*54*)? Previous studies on memory suppression in healthy individuals provide some arguments in favor of the existence of a preexisting factor to combat or adequately resist the stress induced by traumatic reviviscence. Individuals with better engagement of the control system experience fewer memory intrusions (*34*, *36*), greater disruption of perceptual memory (*27*), and greater forgetting (*25*, *26*, *28*–*30*, *36*, *37*). Lower attentional control capacities (*55*) or deficient retrieval suppression (*56*) are potential risk factors for the development of intrusive memories after emotional films.

Memory control mechanisms may also adapt after exposure to stressful events to overcome traumatic experiences (*53*), illustrating a form of acquired resilience. The stronger top-down suppression of the ventral precuneus observed in trauma-exposed individuals without PTSD compared not only with individuals with PTSD but also nonexposed individuals is interesting in that respect. The precuneus seems central to the representation of sensory and mental images of the trauma (*57*–*59*), disconnected from contextual representations in the hippocampus (*1*). Suppression of the precuneus is compatible with recent findings suggesting that new memory engrams can be rapidly encoded (*60*) and updated (*61*) into this region. The coordinate suppression of intrusive memories across the precuneus and hippocampus, which we observed specifically in resilient individuals,

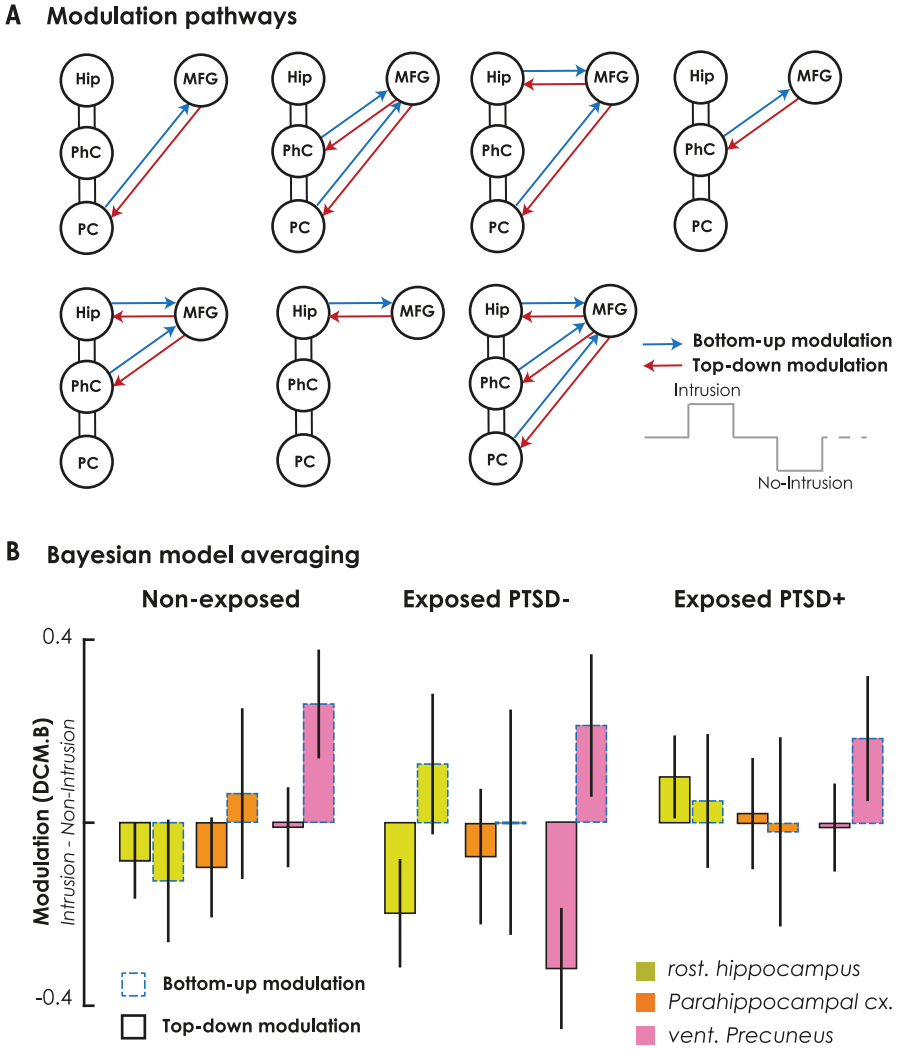


Fig. 5. DCM model space and coupling parameters. (**A**) Bottom-up and top-down influences between the right anterior MFG and memory regions during suppression attempts were measured across seven DCM models capturing different connection pathways. The modulatory input acting on these connections reflected the difference in coupling between intrusive and nonintrusive memories. Memory target regions included rostral hippocampus (Hip), parahippocampal cortex (PhC), and precuneus (PC). (**B**) Bayesian model averaging across model space of the top-down and bottom-up modulatory parameters. Error bars reflect 95% confidence intervals and indicate significance when they do not encompass zero.

might therefore be crucial to cope with traumatic events.

The disruption of memory control mechanisms seen in PTSD might also reflect a form of acquired vulnerability in PTSD or a preexisting vulnerability of inhibitory mechanisms. Stress can impair executive functioning (*62*), including cognitive control (*63*). Animal models propose that excessive and repeated stress damages GABAergic interneurons in the hippocampus (*64*), a neurotransmitter which potentially mediates the inhibitory effect associated with memory suppression (*29*, *41*) and whose receptor population is disrupted after trauma (*65*). Similarly, an alteration of the white-matter tracts that propagate the inhibitory command (*66*) could also prevent this effect from taking place in individuals with PTSD.

Treating mechanisms of suppression?

The cross-sectional study described here does not provide insight into the origin of the observed memory suppression deficit seen in PTSD. However, it provides important information concerning the role of memory suppression mechanisms for understanding and treating the development of PTSD. Most of the current recommended psychotherapeutic treatments for PTSD focus on the traumatic experience and involve, to some degree, a reexposure to the traumatic content, which can sometimes be problematic in clinical settings (*10*). Treatments focused on the memory control system, using neutral material unrelated to the trauma, might also be a viable option to complement standard psychological interventions and help patients to gain a better control over their memories during therapy. The capacity to benefit from exposure therapy in PTSD depends on prefrontal control resources (*67*, *68*) and on the propagation of neural flows originating from the right anterior DLPFC (*69*).

However, the effectiveness of a treatment may be limited if applied in the context of compromised capacity and impaired functional brain connectivity. Nonetheless, individuals with PTSD have shown some residual capacities. Analysis of local activity revealed that these individuals could still engage the memory control network during attempts to suppress memories, although this did not translate into a reduction of intrusion frequency. Analysis of connectivity also revealed preserved suppression processes in PTSD when memory cues failed to trigger intrusion. In fact, PTSD might excessively rely on proactive control (*70*), an anticipatory process attempting to gate memory retrieval before intrusion arises to conscious awareness. Excessive proactive control could reduce the opportunities to modulate intrusive memory traces and lead to the same paradoxical and harmful avoidance effect on traumatic memory. Suppression can also induce forgetting of contextual information associated with the reminder cue (*38*). In the context of PTSD, exaggerated anticipatory suppression could therefore prevent the learning of safe contextual cues and promote overgeneralization of fear. Interventions focused on training the memory control system should aim for better allocation of the preserved resources of the control system and proactive engagement.

It remains unknown whether the mechanisms identified here can disrupt the traumatic memory itself, as trauma-focused exposure treatments can. Suppression can be ineffective after consolidation (*71*) or when memory reactivation is too strong (*72*). Suppression can also be detrimental to emotional response if individuals show poor inhibitory capacities or when forgetting is impossible (*34*, *35*). Suppressing traumatic memory should thus not be attempted in individuals while they lack the necessary coping skills of inhibition and intrusive memories remain vivid and salient. Once these coping skills are strengthened, and traumatic traces have been reprocessed by the hippocampus together with contextual representations during standard exposure therapy sessions (*15*), remediation of control capacity might also promote the disruption and updating of the traumatic engram.

Our findings suggest that the general mental operations usually engaged to banish and suppress the intrusive expression of unwanted memories might contribute to positive adaptation in the aftermath of a traumatic event,

paving the way for new treatments unrelated to the trauma and promoting resilience (*54*).

Materials and methods

Participants

Eighty nonexposed and 120 exposed subjects participated in this study. Exposed participants were recruited through a transdisciplinary and longitudinal research "Programme 13-Novembre" (www.memoire13novembre.fr/), a nationwide funded program supported by victims' associations. Data from seven nonexposed participants were excluded from further analyses for the following reasons: absence of intrusion rating owing to technical or behavioral issues (n = 4), artifacts in the MRI images (n = 2), and inability to pursue the experiment (n = 1). Data from 18 exposed participants were excluded from further analyses for the following reasons: absence of intrusion rating owing to technical or behavioral issues (n = 8), interruption of participation during the MRI acquisition (n = 3), and nonrespect of inclusion criteria (n = 7). Among these seven participants who did not respect the inclusion criteria in the exposed group, six met the criteria for the reexperiencing symptoms but without the presence of other symptom categories (including functional significance, i.e., criterion G), and one was not actually exposed to the attacks (criterion A). The final sample consists of 102 participants exposed to the 13 November 2015 terrorist attacks in Paris and 73 nonexposed healthy control participants. Nonexposed participants were not present in Paris on 13 November 2015 and were recruited from a local panel of volunteers. All participants were between 18 and 60 years old, right-handed, French speaking, and had a body mass index <35 kg/m^2. A clinical interview with a medical doctor was conducted to ensure that participants had no reported history of neurological, medical, visual, memory, or psychiatric disorders. Exclusion criteria also included history of alcohol or substance abuse (other than nicotine), mental or physical conditions that preclude MRI scanning (e.g., claustrophobia or metal implants), and medical treatment that may affect the central nervous system or cognitive functions. Fourteen exposed participants were taking antidepressant, anxiolytic, and/or hypnotic medication at the time of the study (see table S15 for a detailed description of psychoactive medication). We decided to include medicated and unmedicated exposed participants to reflect the general PTSD population. However, additional analyses of covariance were carried out to ensure that the main findings did not depend on these participants.

Exposed participants were diagnosed using the structured clinical interview for *DSM-5* (SCID) (*73*) conducted by a trained psychologist and supervised by a psychiatrist. All exposed participants met *DSM-5* criterion A, indicating that they experienced a traumatic event. Different types of exposure to the Paris attacks were observed in our sample (see table S1). *DSM-5* exposure types include: (i) individuals directly targeted by the terrorist attacks (criterion A1) or (ii) witnessing the attacks (criterion A2); (iii) close relatives of a deceased victim of the attacks (criterion A3); (iv) individuals who were exposed to aversive scenes and the attacks as first responders and police officers. Exposed participants were diagnosed with PTSD in its full form if all the additional diagnostic criteria defined by *DSM-5* were met (n = 29). Participants were diagnosed with PTSD in its partial form (n = 26) if they had reexperiencing symptoms (criterion B), with symptoms persisting for more than one month (criterion F) that caused significant distress and functional impairment (criterion G). For this partial form of PTSD, >80% of the individuals also suffered from two other symptom criteria [i.e., avoidance (C), negative alterations in cognition and mood (D), or hyperarousal (E)]. Subthreshold (also referred to as partial or subsyndromal) PTSD has been associated with clinically significant psychological, social, and functional impairments (*48*). Although participants with a partial PTSD profile did not meet the full clinical symptoms of PTSD, the intrusive symptoms identified in each participant caused important distress that may be associated with significant levels of social and functional impairments comparable to full PTSD (*74*). The concept of subthreshold (partial or subsyndromal) PTSD suggests that an individual may still display noticeable clinical impairment (*75*), especially in relation to reexperiencing and intrusive symptoms, while not meeting full criteria for either avoidance or hyperarousal symptoms (*76*, *77*). Therefore, trauma-exposed participants with full and partial PTSD profiles were grouped together for the purpose of statistical analyses in one clinical group referred to as the PTSD group. The study includes 55 trauma-exposed participants with PTSD (PTSD+), 47 trauma-exposed participants without PTSD (PTSD−), and 73 nonexposed control participants (Control).

PTSD symptom severity was assessed with the Post-traumatic Stress Disorder Checklist for *DSM-5* (PCL-5) (*78*). To assess for anxiety and depression, State-Trait Anxiety Inventory (STAI) (*79*) and Beck Depression Inventory (BDI) (*80*) were also administered. Participants' sleep habits during the month preceding their inclusion in the study were assessed with the Pittsburgh Sleep Quality Index (*81*), and the presence of sleep insomnia was measured with the Insomnia Severity Index. To compare the participants' usual sleep duration with their sleep duration the night before MRI acquisition, we computed an ANOVA with as within-factor the sleep duration (usual and night-before acquisition) and as between-factor the four groups of subjects. We found an effect of sleep duration [$F_{1,158}$ = 13.43, P < 0.001] with no interaction with the group [$F_{3,158}$ = 0.02, P = 0.996] that indicated a decreased sleep duration the night before the acquisition in all participants. Tukey post-hoc comparisons for the group effect showed that the nonexposed group reported longer sleep duration than the participants with complete (P = 0.03) and incomplete (P = 0.013) PTSD. However, no differences were observed among the groups of exposed participants (P > 0.3). The demographic and clinical characteristics of participants are summarized in table S1.

All participants completed the study between 13 June 2016 and 7 June 2017. The exposed groups did not significantly differ in the delay between the date of the Paris attacks and the date of inclusion in the study ($F_{2,99}$ = 2.06, P = 0.13; PTSD absent = 1.14 ± 0.18 years, partial PTSD = 1.23 ± 0.21 years, full PTSD = 1.14 ± 0.23 years). Participants were financially compensated for their participation in the study. The study was approved by the regional research ethics committee (Comité de Protection des Personnes Nord-Ouest III, sponsor ID: C16-13, RCB ID: 2016-A00661-50, clinicaltrials.gov registration number: NCT02810197). All participants gave written informed consent before participation, in agreement with French ethical guidelines. Participants were asked not to consume psychostimulants, drugs, or alcohol before or during the experimental period.

Materials

The stimuli were three series of lists of 72 word-object pairs composed of neutral abstract French words (*82*) and objects selected from the Bank Of Standardized Stimuli (BOSS) (*83*). Three series of four lists of 18 pairs assigned to four conditions (think, no-think, baseline, and unprimed) were created, plus eight fillers used for practice. The lists of pairs were presented in counterbalanced order across the three series, the four conditions and the three groups of participants and matched on different properties that may influence performance to the task. The lists of words were matched on average naming latency, number of letters, and lexical frequency (*82*). The lists of objects were matched relative to the naming latency, familiarity and visual complexity levels, viewpoint, name and object agreement, and manipulability (*83*). Stimuli were presented using the Psychophysics Toolbox implemented in MATLAB (MathWorks). We used neutral material completely disconnected from the traumatic experience, which enabled the investigation of general memory control mechanisms and

incidentally avoided ethical issues for the trauma-exposed group.

Procedure

Before MRI acquisition, participants learned 72 French neutral word-object pairs that were presented for 5 s each. After the presentation of all pairs, the word cue for a given pair was presented on the screen for up to 4 s, and participants were asked whether they could recall and fully visualize the paired object (see Fig. 1B for details of the procedure). If so, three objects then appeared on the screen (one correct and two foils), and participants had up to 4 s to select which object was associated with the word cue. After each recognition test, the object correctly associated with the word appeared for 2500 ms on the screen, and participants were asked to use this feedback to increase their knowledge of the pair. Pairs were learned through this test–feedback cycle procedure until either the learning criterion (at least 90% correct responses) was reached or a maximum of six presentations was achieved. Once participants had reached the learning criterion, their memory was assessed one last time using a final criterion test on all of the pairs but without giving any feedback on the response. Note that no differences were found between groups on this final criterion test (all $P > 0.18$), suggesting that our procedure carefully matched the learning of word-object associations between groups. After this learning phase, pairs were divided into three lists of 18 pairs assigned to think, no-think, and baseline conditions for the think/no-think task (TNT). Participants were given the think/no-think phase instructions and a short TNT practice session before MRI acquisition to familiarize them to the task.

After this TNT practice session, participants entered the MRI scanner. During the T1 structural image acquisition, the complete list of learned pairs was presented once again to reinforce the learning of the pairs (5 s for each pair). This overtraining procedure was intended to ensure that the word cue would automatically bring back the associated object, allowing us to isolate brain regions engaged to control the intrusion of the paired object during the TNT phase. After this reminder of the pairs, participants performed the TNT task, which was divided into four sessions of ~8 min each. In each session, the 18 think and 18 no-think items were presented twice. Word cues appeared for 3 s on the screen and were written either in green for think trials or in red for no-think trials. During the TNT practice session, participants were trained to use a direct suppression strategy. During the think trials, participants were told to imagine the associated object in as much detail as possible. During the no-think trials, participants were instructed to imperatively prevent the object from coming to mind and to fixate and concentrate on the word cue without looking away. Participants were asked to block thoughts of the object by blanking their mind and not by replacing the object with any other thoughts or mental images. If the object image came to mind anyway, they were asked to push it out of mind. After the end of each of the think or no-think trial cues, participants reported whether the associated object had entered awareness by pressing one of two buttons corresponding to "yes" (i.e., even if the associated object pops very briefly into their mind) or "no." Although participants had up to 3600 ms to make this intrusion rating, they were instructed to make it quickly without thinking and dwelling too much on the associated object. The rating instruction was presented for up to 1 s on the screen and followed by a jittered fixation cross (1400, 1800, 2000, 2200, or 2600 ms). The Genetic Algorithm toolbox (*84*) was used to optimize the efficiency of the think versus no-think contrast. Twenty percent additional null events with no duration and followed by the jittered fixation cross only were added.

The perceptual identification task followed the TNT phase and tested whether previous attempts at suppression affected repetition priming. It comprised a single session of about 8 min. Each think, no-think, baseline, and unprimed item was presented on one trial in a 500 pixel by 500 pixel frame centered on a gray background, and trials were separated by a fixation cross. During each trial, a single item was gradually presented using a phase-unscrambling procedure that lasted for 3.15 s. Participants' instruction was to watch carefully as the object was progressively unscrambled and to press the button as fast as possible when they were able to see and name the object in the picture. Unscrambling continued until a complete image appeared, irrespective of when and whether participants pressed a button. The scrambling was achieved by decomposing the picture into phase and amplitude spectra using a Fourier transform. Random noise was added to the phase spectrum starting from 100% and was decreased by 5% steps until 0% (i.e., intact picture) was reached. The picture was presented at each level of noise for 150 ms, yielding a total stimulus duration of 3.15 s. Between trials, there was a 2.4-s average interstimuli interval, and there were also 20% additional null events added. Brain activity was also recorded during this perceptual identification task but data are not reported here. After this task, a resting-state recording was also proposed to the participants. During this session, participants were instructed to keep their eyes closed, to let their thoughts flow freely without focusing on any particular idea, and to remain still and awake.

Finally, during a debriefing questionnaire, participants were asked about the strategies used during the TNT phase. Participants rated on a five-point scale [never (0) to all the time (4)] the degree to which they used different kind of strategies to prevent the object from coming to mind during the no-think condition (i.e., direct suppression, thought substitution, or another strategy). This questionnaire was administered to determine whether participants complied with the direct suppression instructions. Debriefing confirmed that the participants remained attentive to the word displayed on the screen and predominantly controlled the unwanted memories by directly suppressing the associated object. Participants engaged significantly less in other strategies than in direct suppression to control awareness of the no-think items (Wilcoxon signed-rank test: $z > 140$, $P < 0.001$). Moreover, Kruskal-Wallis tests did not evidence any difference between the groups for any kind of strategies used [$H(2) < 2.73$, $P > 0.26$]. The mean rating score for each strategy is displayed in fig. S4 for each group.

MRI acquisition parameters

MRI data were acquired on a 3T Achieva scanner (Philips). All participants first underwent a high-resolution T1-weighted anatomical volume imaging using a 3D fast field echo (FFE) sequence (3D-T1-FFE sagittal; TR = 20 ms, TE = 4.6 ms, flip angle = 10°, SENSE factor = 2, 180 slices, 1 mm by 1 mm by 1 mm voxels, no gap, FoV = 256 mm by 256 mm by 180 mm, matrix = 256 by 130 by 180). This acquisition was followed by the TNT functional sessions and an eyes-closed resting-state fMRI sequence, which were acquired using an ascending T2-star EPI sequence (MS-T2-star-FFE-EPI axial; TR = 2050 ms, TE = 30 ms, flip angle = 78°, 32 slices, slice thickness = 3 mm, 0.75-mm gap, matrix 64 by 64 by 32, FoV = 192 mm by 192 mm by 119 mm, 235 volumes per run). Each of the TNT and resting-state functional sequence lasted about 8 min.

fMRI preprocessing

Image preprocessing was first conducted with the Statistical Parametric Mapping software (SPM 12, University College London, London, UK). Functional images were (i) spatially realigned to correct for motion (using a six-parameter rigid body transformation); (ii) corrected for slice acquisition temporal delay; and (iii) co-registered with the skull-stripped structural T1 image. The T1 image was bias-corrected and segmented using tissue probability maps for gray matter, white matter, and cerebrospinal fluid. The forward deformation field (y_*.nii) was derived from the nonlinear normalization of individual gray matter T1 images to the T1 template of the Montreal Neurological Institute (MNI). Each point in this deformation field is a mapping between MNI standard space to native-space

coordinates in millimeters. Thus, this mapping was used to project the coordinates of the MNI standard space ROIs to the native space functional images. All subsequent analyses were conducted using these projected native space ROIs without any spatial warping nor smoothing of the functional images.

Think/no-think univariate analyses

The preprocessed fMRI time series at each voxel were high-pass filtered using a cutoff period of 128 s. Task-related regressors within a GLM for each ROI were created by convolving a boxcar function at stimulus onset for each condition of interest (i.e., think, intrusion, and nonintrusion) with the canonical hemodynamic response function (HRF). Additional regressors of no interest were the six realignment parameters to account for linear residual motion artifacts and session dummy regressors. Filler items, along with the few items with no button press or not correctly recalled during think condition, were also entered into a single regressor of no interest. Autocorrelation between fMRI time series was corrected using a first-order autoregressive AR(1) model of noise temporal autocorrelation and the GLM parameters were estimated using restricted maximum likelihood (ReML). Voxel-based analyses were performed by entering first-level activation maps for each condition of interest into flexible ANOVAs implemented in SPM, which used pooled error and correction for nonsphericity to create *t*-statistics. The SPMs were thresholded for voxels whose statistic exceeded a peak threshold corresponding to $P_{FWE} < 0.05$ family-wise error (FWE) correction using random field theory across the whole brain (for the no-think versus think contrasts), or within the appropriate search volumes of interest to perform within- and between-group comparisons for the intrusion versus nonintrusion contrasts (using an initial threshold of $P_{uncorr} < 0.005$). Additional exploratory analyses were performed to examine the relation between brain activation (intrusion > nonintrusion) and intrusion frequency using a separate regression model for each group of participants ($P_{uncorr} < 0.005$).

Regions of interest (ROIs)

We focused on prefrontal and memory systems previously identified in the TNT literature as up-regulated and down-regulated, respectively, during the attempts to suppress unwanted memories. We selected ROIs from the Brainnetome atlas (*85*; http://atlas.brainnetome.org/) that overlap with these control and memory networks. The Brainnetome atlas is a fine-grained connectivity-based and cross-validated parcellation atlas of the brain into 210 cortical and 36 subcortical regions and is therefore ideally suited to study the change in task-based connectivity across the control and memory networks. Given the strong right lateralization of the prefrontal control network during memory inhibition, we selected brain regions of the right hemisphere consistently activated during memory retrieval suppression (*25*–*30*, *34*, *36*, *37*), including: (i) the right superior frontal gyrus (SFG); (ii) the core of the right middle frontal gyrus (MFG), excluding the posterior sensory-motor inferior frontal junction (center coordinates: $x = 42$, $y = 11$, $z = 39$), as well as the anterior lateral area corresponding to Brodmann area (BA) 10 (center coordinates: $x = 25$, $y = 61$, $z = -4$); (iii) the right inferior frontal gyrus (IFG); and (iv) the right anterior cingulate gyrus (CG). For the memory network, we selected bilateral brain regions consistently reported as suppressed during memory suppression (*25*–*30*, *34*, *36*, *37*), including: (i) the hippocampus (divided into rostral and caudal parts); (ii) the parahippocampal gyrus; (iii) the fusiform gyrus; and (iv) the ventral part of the precuneus alongside the parietal sulcus. The ventral part of the precuneus is associated with visual imagery (*86*), episodic (*60*), autobiographical (*87*), and trauma-related memories (*57*, *58*). Note that the dorsal portion of the precuneus, as well as the transitional zone (BA 31) are activated rather than suppressed during no-think trials, and therefore cannot be included in the down-regulated target memory network. The individual connectivity matrices were estimated on the basis of the prefrontal control network ROIs that comprised 20 regions and the memory networks that included 18 potential sites of suppression (see table S8 for a list of the Brainnetome regions with their labels and center coordinates). For between-group comparisons during connectivity analyses (PPI and DCM), we used the anterior portion of the right MFG (area 46 and ventral area 9/46 of the Brainnetome atlas; see table S8).

Functional connectivity analysis

The regional BOLD signal that was filtered, whitened, and adjusted for confounds was used to perform psychophysiological interaction (PPI) analyses (*51*). We adapted the generalized form of context-dependent PPI (*51*) to investigate task-induced functional connectivity between ROIs of the prefrontal control (i.e., seed) and memory (i.e., target) networks (see table S8), focusing on the contrast involving the suppression of intrusive and nonintrusive memories. Our design optimize the detection of signal change between conditions by imposing short interstimuli intervals and slow changes between main conditions (*88*–*90*). In an attempt to reduce the duration of the task for the sake of the participants, periods of recording without stimulation were scarce and short. This approach, however, prevents the estimation of absolute change in task-induced changes relative to implicit rest baseline (the intercept of the GLM which captures the mean of the signal left unexplained). Moreover, rest baseline in such design are likely contaminated by task-based cognitive processes, which presumably do not abruptly terminate at the onset of resting periods. As such, quantification of absolute change in task-based connectivity is problematic and a contrast approach is usually recommended. To circumvent this problem, we additionally used a blind-deconvolution approach to detect spontaneous event-related changes (*52*) in the resting-state signal of a sequence collected after the TNT task. Onsets of pseudo-events during resting state were obtained for each ROI from BOLD activation using a threshold between 1 and 4 standard deviations from the mean. Once identified, a GLM was estimated for each ROI over all possible micro-time onsets of the neural stick function that could have generated these pseudo-events. We allowed a 3- to 9-s shift to find the best explaining onset of BOLD activation peaks based on the residuals of the GLM.

BOLD time-courses in each seed ROI for both TNT and resting-state sequences were deconvolved to estimate the neural activity. A full-rank cosine basis set convolved with the HRF, as well as the filtered and whitened matrix of confounds, was used as the design matrix of a hierarchical linear model to estimate the underlying neuronal activity under a parametric empirical Bayes scheme (*91*). PPI regressors were created by multiplying estimated neural activity with a boxcar function (modeled as a 3-s short-epoch) encoding TNT or resting-state events. This interaction term was subsequently reconvolved with the canonical HRF and resample to scan resolution. PPI regressors were detrended and normalized to unit length using their norm to facilitate comparisons between TNT and resting-state estimates of connectivity. For each TNT and resting-state sequence, a first-level GLM was created to estimate the connectivity between seed and target preprocessed time-series (data filtered, whitened, and adjusted for confounds). This GLM included in the design matrix the PPI regressors of the seed, the psychological regressors obtained from the convolution of stimulus boxcar function with HRF to control for task-evoked univariate changes, the physiological BOLD signal of the seed region, and a constant term.

Effective connectivity analyses

DCM explains changes in regional activity in terms of experimentally defined modulations ("modulatory input") of the connectivity between regions. Here, we used DCM and Bayesian Model Averaging (BMA; *92*) to assess, in each of our group, whether the modulation in connectivity between the right anterior MFG and memory systems arising from the elevated

control demand during the suppression of intrusive memories (compared with nonintrusions) was mediated by a top-down process.

DCM entails defining a network of a few ROIs and the forward and backward connections between them. The neural dynamics within this network are based on a set of simple differential equations (the bilinear state equation was used here) relating the activity in each region to (i) the activity of other regions via intrinsic connections relative to implicit unmodelled baseline, (ii) experimentally defined extrinsic input (or "driving input") to one or more of the regions, and, most importantly, (iii) experimentally defined modulations (or "modulatory input") in the connectivity between regions. Changes in the network dynamics are caused by these driving (entering-regions) or modulatory (between-regions) inputs. These neural dynamics are then mapped to the fMRI time series using a biophysical model of the BOLD response. The neural (and hemodynamic) parameters of this DCM are estimated using approximate variational Bayesian techniques to maximize the free-energy bound on the Bayesian model evidence. Here, we defined different models defining potential pathways of both top-down and bottom-up modulation between the right MFG and memory systems, and we used BMA to marginalize over these models to derive posterior densities on model parameters that account for model uncertainty.

Retrieval inhibition was assumed to originate from the anterior portion of the right MFG (see ROIs section). Therefore, we focused on the influence of this region over memory regions within the same hemisphere as done in previous studies analyzing effective connectivity using the TNT paradigm (*26*, *27*, *34*, *36*). Note that DCM requires a restricted number of nodes so we focused this analysis on the MTL (including rostral hippocampus and parahippocampal gyrus), as done previously (*26*, *34*, *36*), and on the precuneus for both its functional role in traumatic memories and its strong down-regulation during PPI analyses in healthy participants compared to PTSD group. The caudal hippocampus was not included in this analysis given the absence of significant modulation in this region during PPI analyses. This DCM analysis was conducted on the exact same filtered, whitened, and adjusted for confounds time-series than the ones used for PPI analyses.

Seven DCM models were created (for an illustration of the model space, see Fig. 5A), plus an additional null model. This null model did not include any modulatory input modelling the effect of suppression on connections. This null model was compared to other modulatory models to ensure that suppression induced some modulation of the connections. All models were fully connected and included a common driving input source entering the right MFG and reflecting cue-onset of all trials. The modulatory input acting on intrinsic connections was modeled as a 3-s short-epochs function reflecting the contrast between intrusion and non-intrusion. After estimating all 8 models for each participant (version DCM12.5 revision 7479), we first performed Bayesian model selection (BMS) to compare models including a modulatory input to null model. BMS overwhelmingly favored models including a modulatory input, with an exceedance probability (EP) and expected posterior probability (EPP), of EP = [100% 100% 100%] and EPP = [91% 88% 78%] for nonexposed, non-PTSD, and PTSD groups, respectively.

We then performed BMA including all modulatory models for each group separately. This produces a maximal a posteriori estimate of coupling parameters weighted by the subject specific posterior and by the posterior probability that subject n uses model m, treating the optimal model across participants in each group as a random effect.

Statistical analyses

All a priori hypotheses test of memory suppression-induced changes in functional connectivity were performed using one-sided paired sample t tests for within-group comparisons, and one-sided two-sample t tests for between group comparisons. The expected proportion of type I error across multiple testing was controlled for using the false discovery rate (FDR) correction, with a desired FDR q = 0.05 and assuming a positive dependency between conditions (*93*). In addition, we used a Bayesian approach (*94*) using Markov chain Monte Carlo (MCMC) method. Bayes factors (BF) were estimated for visualization purpose to represent the likelihood of suppression effects for each within-group comparison. Based on this hypothesis, we defined a region of practical equivalence (ROPE) set as a Cohen's d effect size greater than −0.1. The MCMC method generated 50,000 credible parameter combinations that are representative of the posterior distribution. Then, the BF was estimated as the ratio of the proportion of the posterior within the ROPE relative to the proportion of the prior within the ROPE. The conventional interpretation of the magnitude of the BF is that there is substantial evidence for the alternative hypothesis when the BF ranges from 3 to 10, a strong evidence between 10 and 30, a very strong evidence between 30 and 100, and a decisive evidence above 100 (*95*). For ROI analyses, group-level inferences were also conducted using nonparametric random effects statistics to test for within-group differences by bootstrapping the subject set with 5000 iterations and compute 95% confidence intervals. Moreover, group comparisons were also conducted using an ANCOVA model controlling for age, sex, education, medication, duration, and type of exposure to the attacks (table S11). For DCM, BMA gives for each group the mean and standard deviation of the coupling parameters posterior distribution. In line with the DCM Bayesian framework, we estimated the posterior probability and the 95% confidence interval of the within- and between-group differences. In this Bayesian framework, the posterior probability indicates the probability that a random sample from this estimated distribution will be different than zero, and is usually considered as significant when equal to or greater than 0.95 (see also table S14 for an ANCOVA model on individual coupling parameters extracted during BMA).

REFERENCES AND NOTES

1. C. R. Brewin, J. D. Gregory, M. Lipton, N. Burgess, Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychol. Rev.* **117**, 210–232 (2010). doi: 10.1037/a0018113; pmid: 20063969
2. A. Ehlers, D. M. Clark, A cognitive model of posttraumatic stress disorder. *Behav. Res. Ther.* **38**, 319–345 (2000). doi: 10.1016/S0005-7967(99)00123-0; pmid: 10761279
3. C. R. Brewin, The nature and significance of memory disturbance in posttraumatic stress disorder. *Annu. Rev. Clin. Psychol.* **7**, 203–227 (2011). doi: 10.1146/annurev-clinpsy-032210-104544; pmid: 21219190
4. A. Hackmann, A. Ehlers, A. Speckens, D. M. Clark, Characteristics and content of intrusive memories in PTSD and their changes with treatment. *J. Trauma. Stress* **17**, 231–240 (2004). doi: 10.1023/B:JOTS.0000029266.88369.fd; pmid: 15253095
5. A. Ehlers, A. Hackmann, T. Michael, Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory* **12**, 403–415 (2004). doi: 10.1080/09658210444000025; pmid: 15487537
6. J. C. Shipherd, J. G. Beck, The role of thought suppression in posttraumatic stress disorder. *Behav. Ther.* **36**, 277–287 (2005). doi: 10.1016/S0005-7894(05)80076-0
7. A. Nickerson *et al*., Emotional suppression in torture survivors: Relationship to posttraumatic stress symptoms and trauma-related negative affect. *Psychiatry Res.* **242**, 233–239 (2016). doi: 10.1016/j.psychres.2016.05.048; pmid: 27294797
8. C. Purdon, Thought suppression and psychopathology. *Behav. Res. Ther.* **37**, 1029–1054 (1999). doi: 10.1016/S0005-7967(98)00200-9; pmid: 10500319
9. E. B. Foa, T. M. Keane, M. J. Friedman, J. A. Cohen, *Effective Treatments for PTSD: Practice Guidelines from the International Society for Traumatic Stress Studies* (Guilford Press, ed. 2, 2009).
10. C. R. Brewin, Memory and forgetting. *Curr. Psychiatry Rep.* **20**, 87 (2018). doi: 10.1007/s11920-018-0950-7; pmid: 30155780
11. S. Lissek, B. van Meurs, Learning models of PTSD: Theoretical accounts and psychobiological evidence. *Int. J. Psychophysiol.* **98**, 594–605 (2015). doi: 10.1016/j.ijpsycho.2014.11.006; pmid: 25462219
12. I. Liberzon, J. L. Abelson, Context processing and the neurobiology of post-traumatic stress disorder. *Neuron* **92**, 14–30 (2016). doi: 10.1016/j.neuron.2016.09.039; pmid: 27710783
13. R. J. Fenster, L. A. M. Lebois, K. J. Ressler, J. Suh, Brain circuit dysfunction in post-traumatic stress disorder: From mouse to man. *Nat. Rev. Neurosci.* **19**, 535–551 (2018). doi: 10.1038/s41583-018-0039-7; pmid: 30054570
14. S. Kida, Reconsolidation/destabilization, extinction and forgetting of fear memory as therapeutic targets for PTSD. *Psychopharmacology* **236**, 49–57 (2019). doi: 10.1007/s00213-018-5086-2; pmid: 30374892
15. A. Desmedt, A. Marighetto, P.-V. Piazza, Abnormal fear memory as a model for posttraumatic stress disorder. *Biol. Psychiatry* **78**, 290–297 (2015). doi: 10.1016/j.biopsych.2015.06.017; pmid: 26238378
16. J. C. Magee, K. P. Harden, B. A. Teachman, Psychopathology and thought suppression: A quantitative review. *Clin. Psychol. Rev.* **32**, 189–201 (2012). doi: 10.1016/j.cpr.2012.01.001; pmid: 22388007
17. L. S. Bishop, V. E. Ameral, K. M. Palm Reed, The impact of experiential avoidance and event centrality in trauma-related

rumination and posttraumatic stress. *Behav. Modif.* **42**, 815–837 (2018). doi: 10.1177/0145445517747287; pmid: 29241356
18. T. Dalgleish, B. Hauer, W. Kuyken, The mental regulation of autobiographical recollection in the aftermath of trauma. *Curr. Dir. Psychol. Sci.* **17**, 259–263 (2008). doi: 10.1111/j.1467-8721.2008.00586.x
19. M. I. Davies, D. M. Clark, Thought suppression produces a rebound effect with analogue post-traumatic intrusions. *Behav. Res. Ther.* **36**, 571–582 (1998). doi: 10.1016/S0005-7967(98)00051-5; pmid: 9648331
20. J. P. Mitchell *et al*., Separating sustained from transient aspects of cognitive control during thought suppression. *Psychol. Sci.* **18**, 292–297 (2007). doi: 10.1111/j.1467-9280.2007.01891.x; pmid: 17470250
21. D. M. Wegner, Ironic processes of mental control. *Psychol. Rev.* **101**, 34–52 (1994). doi: 10.1037/0033-295X.101.1.34; pmid: 8121959
22. C. Gagne, P. Dayan, S. J. Bishop, When planning to survive goes wrong: Predicting the future and replaying the past in anxiety and PTSD. *Curr. Opin. Behav. Sci.* **24**, 89–95 (2018). doi: 10.1016/j.cobeha.2018.03.013
23. M. B. Stein, M. P. Paulus, Imbalance of approach and avoidance: The yin and yang of anxiety disorders. *Biol. Psychiatry* **66**, 1072–1074 (2009). doi: 10.1016/j.biopsych.2009.09.023; pmid: 19944792
24. M. C. Anderson, C. Green, Suppressing unwanted memories by executive control. *Nature* **410**, 366–369 (2001). doi: 10.1038/35066572; pmid: 11268212
25. B. E. Depue, T. Curran, M. T. Banich, Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science* **317**, 215–219 (2007). doi: 10.1126/science.1139560; pmid: 17626877
26. R. G. Benoit, M. C. Anderson, Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron* **76**, 450–460 (2012). doi: 10.1016/j.neuron.2012.07.025; pmid: 23083745
27. P. Gagnepain, R. N. Henson, M. C. Anderson, Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1310–E1319 (2014). doi: 10.1073/pnas.1311468111; pmid: 24639546
28. B. E. Depue, J. M. Orr, H. R. Smolker, F. Naaz, M. T. Banich, The organization of right prefrontal networks reveals common mechanisms of inhibitory regulation across cognitive, emotional, and motor processes. *Cereb. Cortex* **26**, 1634–1646 (2016). doi: 10.1093/cercor/bhu324; pmid: 25601236
29. T. W. Schmitz, M. M. Correia, C. S. Ferreira, A. P. Prescot, M. C. Anderson, Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nat. Commun.* **8**, 1311 (2017). doi: 10.1038/s41467-017-00956-z; pmid: 29101315
30. M. C. Anderson *et al*., Neural systems underlying the suppression of unwanted memories. *Science* **303**, 232–235 (2004). doi: 10.1126/science.1089504; pmid: 14716015
31. M. C. Anderson, S. Hanslmayr, Neural mechanisms of motivated forgetting. *Trends Cogn. Sci.* **18**, 279–292 (2014). doi: 10.1016/j.tics.2014.03.002; pmid: 24747000
32. Y. Wang, A. Luppi, J. Fawcett, M. C. Anderson, Reconsidering unconscious persistence: Suppressing unwanted memories reduces their indirect expression in later thoughts. *Cognition* **187**, 78–94 (2019). doi: 10.1016/j.cognition.2019.02.016; pmid: 30852261
33. K. Kim, D.-J. Yi, Out of mind, out of sight: Perceptual consequences of memory suppression. *Psychol. Sci.* **24**, 569–574 (2013). doi: 10.1177/0956797612457577; pmid: 23459870
34. P. Gagnepain, J. Hulbert, M. C. Anderson, Parallel regulation of memory and emotion supports the suppression of intrusive memories. *J. Neurosci.* **37**, 6423–6441 (2017). doi: 10.1523/JNEUROSCI.2732-16.2017; pmid: 28559378
35. N. Legrand *et al*., Does the heart forget? Modulation of cardiac activity induced by inhibitory control over emotional memories. bioRxiv 376954 [Preprint]. 13 February 2019. https://www.biorxiv.org/content/10.1101/376954v3.
36. R. G. Benoit, J. C. Hulbert, E. Huddleston, M. C. Anderson, Adaptive top-down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *J. Cogn. Neurosci.* **27**, 96–111 (2015). doi: 10.1162/jocn_a_00696; pmid: 25100219
37. B. J. Levy, M. C. Anderson, Purging of memories from conscious awareness tracked in the human brain. *J. Neurosci.* **32**, 16785–16794 (2012). doi: 10.1523/JNEUROSCI.2640-12.2012; pmid: 23175832
38. J. C. Hulbert, R. N. Henson, M. C. Anderson, Inducing amnesia through systemic suppression. *Nat. Commun.* **7**, 11003 (2016). doi: 10.1038/ncomms11003; pmid: 26977589
39. X. Hu, Z. M. Bergström, P. Gagnepain, M. C. Anderson, Suppressing unwanted memories reduces their unintended influences. *Curr. Dir. Psychol. Sci.* **26**, 197–206 (2017). doi: 10.1177/0963721417689881; pmid: 28458471
40. B. E. Depue, A neuroanatomical model of prefrontal inhibitory modulation of memory retrieval. *Neurosci. Biobehav. Rev.* **36**, 1382–1399 (2012). doi: 10.1016/j.neubiorev.2012.02.012; pmid: 22374224
41. M. C. Anderson, J. G. Bunce, H. Barbas, Prefrontal-hippocampal pathways underlying inhibitory control over memory. *Neurobiol. Learn. Mem.* **134**, 145–161 (2016). doi: 10.1016/j.nlm.2015.11.008; pmid: 26642918
42. S. J.-H. van Rooij, E. Geuze, M. Kennis, A. R. Rademaker, M. Vink, Neural correlates of inhibition and contextual cue processing related to treatment response in PTSD. *Neuropsychopharmacology* **40**, 667–675 (2015). doi: 10.1038/npp.2014.220; pmid: 25154707
43. N. Fani *et al*., Attentional control abnormalities in posttraumatic stress disorder: Functional, behavioral, and structural correlates. *J. Affect. Disord.* **253**, 343–351 (2019). doi: 10.1016/j.jad.2019.04.098; pmid: 31078834
44. A. Catarino, C. S. Küpper, A. Werner-Seidler, T. Dalgleish, M. C. Anderson, Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychol. Sci.* **26**, 604–616 (2015). doi: 10.1177/0956797615569889; pmid: 25847536
45. G. T. Waldhauser *et al*., The neural dynamics of deficient memory control in heavily traumatized refugees. *Sci. Rep.* **8**, 13132 (2018). doi: 10.1038/s41598-018-31400-x; pmid: 30177846
46. R. L. Aupperle, A. J. Melrose, M. B. Stein, M. P. Paulus, Executive function and PTSD: Disengaging from trauma. *Neuropharmacology* **62**, 686–694 (2012). doi: 10.1016/j.neuropharm.2011.02.008; pmid: 21349277
47. J. DeGutis *et al*., Posttraumatic psychological symptoms are associated with reduced inhibitory control, not general executive dysfunction. *J. Int. Neuropsychol. Soc.* **21**, 342–352 (2015). doi: 10.1017/S1355617715000235; pmid: 26126233
48. M. Brancu *et al*., Subthreshold posttraumatic stress disorder: A meta-analytic review of *DSM-IV* prevalence and a proposed *DSM-5* approach to measurement. *Psychol. Trauma* **8**, 222–232 (2016). doi: 10.1037/tra0000078; pmid: 26390108
49. D. R. Sullivan *et al*., Behavioral and neural correlates of memory suppression in PTSD. *J. Psychiatr. Res.* **112**, 30–37 (2019). doi: 10.1016/j.jpsychires.2019.02.015; pmid: 30844595
50. I. K. Lyoo *et al*., The neurobiological role of the dorsolateral prefrontal cortex in recovery from trauma. Longitudinal brain imaging study among survivors of the South Korean subway disaster. *Arch. Gen. Psychiatry* **68**, 701–713 (2011). doi: 10.1001/archgenpsychiatry.2011.70; pmid: 21727254
51. D. G. McLaren, M. L. Ries, G. Xu, S. C. Johnson, A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage* **61**, 1277–1286 (2012). doi: 10.1016/j.neuroimage.2012.03.068; pmid: 22484411
52. G. R. Wu *et al*., A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Med. Image Anal.* **17**, 365–374 (2013). doi: 10.1016/j.media.2013.01.003; pmid: 23422254
53. J. C. Hulbert, M. C. Anderson, What doesn't kill you makes you stronger: Psychological trauma and its relationship to enhanced memory control. *J. Exp. Psychol. Gen.* **147**, 1931–1949 (2018). doi: 10.1037/xge0000461; pmid: 30024184
54. R. Kalisch *et al*., The resilience framework as a strategy to combat stress-related disorders. *Nat. Hum. Behav.* **1**, 784–790 (2017). doi: 10.1038/s41562-017-0200-8; pmid: 31024125
55. J. Verwoerd, P. J. de Jong, I. Wessel, Low attentional control and the development of intrusive memories following a laboratory stressor. *J. Psychopathol. Behav. Assess.* **30**, 291–297 (2008). doi: 10.1007/s10862-008-9080-6
56. M. Streb, A. Mecklinger, M. C. Anderson, L. H. Johanna, T. Michael, Memory control ability modulates intrusive memories after analogue trauma. *J. Affect. Disord.* **192**, 134–142 (2016). doi: 10.1016/j.jad.2015.12.032; pmid: 26724692
57. G. Sartory *et al*., In search of the trauma memory: A meta-analysis of functional neuroimaging studies of symptom provocation in posttraumatic stress disorder (PTSD). *PLOS ONE* **8**, e58150 (2013). doi: 10.1371/journal.pone.0058150; pmid: 23536785
58. G. A. Gvozdanovic, P. Stämpfli, E. Seifritz, B. Rasch, Neural correlates of experimental trauma memory retrieval. *Hum. Brain Mapp.* **38**, 3592–3602 (2017). pmid: 28419641
59. I. A. Clark, C. E. Mackay, Mental imagery and post-traumatic stress disorder: A neuroimaging and experimental psychopathology approach to intrusive memories of trauma. *Front. Psychiatry* **6**, 104 (2015). doi: 10.3389/fpsyt.2015.00104; pmid: 26257660
60. S. Brodt *et al*., Fast track to the neocortex: A memory engram in the posterior parietal cortex. *Science* **362**, 1045–1048 (2018). doi: 10.1126/science.aau2528; pmid: 30498125
61. P. L. St. Jacques, C. Olm, D. L. Schacter, Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19671–19678 (2013). doi: 10.1073/pnas.1319630110; pmid: 24191059
62. G. S. Shields, M. A. Sazma, A. P. Yonelinas, The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neurosci. Biobehav. Rev.* **68**, 651–668 (2016). doi: 10.1016/j.neubiorev.2016.06.038; pmid: 27371161
63. S. A. Gagnon, A. D. Wagner, Acute stress and episodic memory retrieval: Neurobiological mechanisms and behavioral consequences. *Ann. N. Y. Acad. Sci.* **1369**, 55–75 (2016). doi: 10.1111/nyas.12996; pmid: 26799371
64. B. Czéh *et al*., Chronic stress reduces the number of GABAergic interneurons in the adult rat hippocampus, dorsal-ventral and region-specific differences. *Hippocampus* **25**, 393–405 (2015). doi: 10.1002/hipo.22382; pmid: 25331166
65. I. Reuveni *et al*., Altered cerebral benzodiazepine receptor binding in post-traumatic stress disorder. *Transl. Psychiatry* **8**, 206 (2018). doi: 10.1038/s41398-018-0257-9; pmid: 30287828
66. S. Siehl, J. A. King, N. Burgess, H. Flor, F. Nees, Structural white matter changes in adults and children with posttraumatic stress disorder: A systematic review and meta-analysis. *Neuroimage Clin.* **19**, 581–598 (2018). doi: 10.1016/j.nicl.2018.05.013; pmid: 29984166
67. G. A. Fonzo *et al*., PTSD psychotherapy outcome predicted by brain activation during emotional reactivity and regulation. *Am. J. Psychiatry* **174**, 1163–1174 (2017). doi: 10.1176/appi.ajp.2017.16091072; pmid: 28715908
68. E. Falconer, A. Allen, K. L. Felmingham, L. M. Williams, R. A. Bryant, Inhibitory neural activity predicts response to cognitive-behavioral therapy for posttraumatic stress disorder. *J. Clin. Psychiatry* **74**, 895–901 (2013). doi: 10.4088/JCP.12m08020; pmid: 24107763
69. A. Etkin *et al*., Using fMRI connectivity to define a treatment-resistant form of post-traumatic stress disorder. *Sci. Transl. Med.* **11**, eaal3236 (2019). doi: 10.1126/scitranslmed.aal3236; pmid: 30944165
70. T. S. Braver, The variable nature of cognitive control: A dual mechanisms framework. *Trends Cogn. Sci.* **16**, 106–113 (2012). doi: 10.1016/j.tics.2011.12.010; pmid: 22245618
71. Y. Liu *et al*., Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. *Nat. Commun.* **7**, 13375 (2016). doi: 10.1038/ncomms13375; pmid: 27898050
72. G. J. Detre, A. Natarajan, S. J. Gershman, K. A. Norman, Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia* **51**, 2371–2388 (2013). doi: 10.1016/j.neuropsychologia.2013.02.017; pmid: 23499722
73. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, ed. 5, 2013).
74. C. Zlotnick, C. L. Franklin, M. Zimmerman, Does "subthreshold" posttraumatic stress disorder have any clinical relevance? *Compr. Psychiatry* **43**, 413–419 (2002). doi: 10.1053/comp.2002.35900; pmid: 12439826
75. N. P. Mota *et al*., High burden of subthreshold DSM-5 post-traumatic stress disorder in U.S. military veterans. *World Psychiatry* **15**, 185–186 (2016). doi: 10.1002/wps.20313; pmid: 27265715
76. E. B. Blanchard *et al*., One-year prospective follow-up of motor vehicle accident victims. *Behav. Res. Ther.* **34**, 775–786 (1996). doi: 10.1016/0005-7967(96)00038-1; pmid: 8952120
77. R. H. Pietrzak *et al*., The burden of full and subsyndromal posttraumatic stress disorder among police involved in the World Trade Center rescue and recovery effort. *J. Psychiatr. Res.* **46**, 835–842 (2012). doi: 10.1016/j.jpsychires.2012.03.011; pmid: 22464942
78. C. A. Blevins, F. W. Weathers, M. T. Davis, T. K. Witte, J. L. Domino, The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): development and initial psychometric evaluation. *J. Trauma. Stress* **28**, 489–498 (2015). doi: 10.1002/jts.22059; pmid: 26606250
79. C. D. Spielberger, M. Bruchon-Schweitzer, I. Paulhan, *Inventaire d'Anxiété état-trait: forme Y* (ECPA, les Éditions du centre de Psychologie Appliquée, 1993).
80. A. T. Beck, R. A. Steer, G. K. Brown, *Manual for the Beck Depression Inventory-II* (Psychological Corporation, ed. 2, 1996).

81. D. J. Buysse, C. F. Reynolds 3rd, T. H. Monk, S. R. Berman, D. J. Kupfer, The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989). doi: 10.1016/0165-1781(89)90047-4; pmid: 2748771
82. A. Syssau, N. Font, Évaluations des caractéristiques émotionnelles d'un corpus de 604 mots. *Bull. Psychol.* **477**, 361–367 (2005). doi: 10.3917/bupsy.477.0361
83. M. B. Brodeur, K. Guérard, M. Bouras, Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PLOS ONE* **9**, e106953 (2014). doi: 10.1371/journal.pone.0106953; pmid: 25211489
84. T. D. Wager, T. E. Nichols, Optimization of experimental design in fMRI: A general framework using a genetic algorithm. *Neuroimage* **18**, 293–309 (2003). doi: 10.1016/S1053-8119(02)00046-0; pmid: 12595184
85. L. Fan *et al.*, The Human Brainnetome Atlas: A new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016). doi: 10.1093/cercor/bhw157; pmid: 27230218
86. M. Grol, G. Vingerhoets, R. De Raedt, Mental imagery of positive and neutral memories: A fMRI study comparing field perspective imagery to observer perspective imagery. *Brain Cogn.* **111**, 13–24 (2017). doi: 10.1016/j.bandc.2016.09.014; pmid: 27816776
87. P. L. St. Jacques, K. K. Szpunar, D. L. Schacter, Shifting visual perspective during retrieval shapes autobiographical memories. *Neuroimage* **148**, 103–114 (2017). doi: 10.1016/j.neuroimage.2016.12.028; pmid: 27989780
88. K. J. Friston, E. Zarahn, O. Josephs, R. N. A. Henson, A. M. Dale, Stochastic designs in event-related fMRI. *Neuroimage* **10**, 607–619 (1999). doi: 10.1006/nimg.1999.0498; pmid: 10547338
89. R. M. Birn, R. W. Cox, P. A. Bandettini, Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *Neuroimage* **15**, 252–264 (2002). doi: 10.1006/nimg.2001.0964; pmid: 11771993
90. O. Josephs, R. N. Henson, Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philos. Trans. R. Soc. London Ser. B* **354**, 1215–1228 (1999). doi: 10.1098/rstb.1999.0475; pmid: 10466147
91. D. R. Gitelman, W. D. Penny, J. Ashburner, K. J. Friston, Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *Neuroimage* **19**, 200–207 (2003). doi: 10.1016/S1053-8119(03)00058-2; pmid: 12781739
92. W. D. Penny *et al.*, Comparing families of dynamic causal models. *PLOS Comput. Biol.* **6**, e1000709 (2010). doi: 10.1371/journal.pcbi.1000709; pmid: 20300649
93. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001). doi: 10.1214/aos/1013699998
94. J. K. Kruschke, Bayesian estimation supersedes the *t* test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013). doi: 10.1037/a0029146; pmid: 22774788
95. R. Wetzels *et al.*, Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspect. Psychol. Sci.* **6**, 291–298 (2011). doi: 10.1177/1745691611406923; pmid: 26168519

ACKNOWLEDGMENTS

We thank all participants for volunteering in this study and the associations of victims who have supported this project. We thank the medical doctors, especially M. Mialon and E. Duprey, and the staff at Cyceron (Biomedical Imaging Platform in Caen). We also thank the researchers; psychologists M. Deschamps, P. Billard, B. Marteau, R. Copalle, and C. Becquet; technicians; and administrative staff at U1077 (Caen), at "Programme 13-Novembre" in Paris, at INSERM "Délégation Régionale Nord-Ouest" (Lille), and at INSERM "Pôle Recherche Clinique", especially K. Ammour. **Funding:** This study was funded by the French Commissariat-General for Investment (CGI) via the National Research Agency (ANR) and the "Programme d'investissement pour l'Avenir (PIA)." The study was realized within the framework of "Programme 13-Novembre" (EQUIPEX Matrice) headed by D.P. and F.E. This program is sponsored by the CNRS and INSERM and supported administratively by HESAM Université, bringing together 35 partners (see www.memoire13novembre.fr). A.M. is funded by a 3-year postdoctoral fellowship from the Normandy region. **Author contributions:** J.D., D.P., F.E., and P.G. designed the study. J.D., D.P., F.E., C.K.-P., and P.G. obtained the financial support. A.M., C.P., and T.V. performed the data acquisition. C.M. and F.F. managed and coordinated the research activity planning and execution. F.V. and V.d.l.S. supervised MRI data collection on human participants and medical interviews. V.d.l.S. supervised the medical aspects of the study, and J.D. supervised SCID interviews and psychiatric examinations. A.M. and P.G. analyzed the behavioral and functional data with the help of G.L. and C.P. A.M. and P.G. wrote the original draft. All authors reviewed and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data and code that support the conclusions of this study are available in the main text and the supplementary material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6479/eaay8477/suppl/DC1
Figs. S1 to S5
Tables S1 to S15

View/request a protocol for this paper from *Bio-protocol*.

25 July 2019; accepted 12 December 2019
10.1126/science.aay8477

Abstract

An opened fundamental question in clinical neuroscience is why following a traumatic experience some individuals can cope with the trauma, while others remain traumatized. The persistence of vivid and stressful intrusive memories is a central clinical feature of post-traumatic stress disorder (PTSD). This psychiatric condition has long been characterized as a memory disorder rooted in the alteration of the hippocampus. Recent studies have proposed that PTSD may be also linked to a general dysfunction of the brain networks supporting the suppression of intrusive memories. The aim of this thesis was to identify the brain markers able to discriminate and predict resilient and maladaptive outcomes following the Paris November 13th terrorist attacks. In the first study, we used computational modelling and brain connectivity analyses to characterize two different brain mechanisms underlying the control of intrusive memories, namely predictive and reactive control. We found that individuals with PTSD showed aberrant beliefs about upcoming intrusive memories, accompanied by exaggerated efforts to prevent them and the parallel incapacity to purge away unwanted memories intruding consciousness. The imbalance between predictive and reactive memory control was related to avoidance symptoms severity. In a second, longitudinal study, we explored how memory control dynamics and hippocampal volumes evolved in individuals remitted from PTSD and individuals with persisting PTSD. We found that, three years after the trauma, the remission from PTSD was associated with plastic hippocampal changes and the recovery of the balance between predictive and reactive control of intrusive memories. These two markers were also predictive of future decrease in PTSD symptoms severity five years after the trauma, revealing that neurocognitive plasticity in control and memory circuits can predict the remission and the persistence of PTSD in time.

Résumé

Une question encore ouverte en neurosciences cliniques est pourquoi certains individus peuvent surmonter une expérience traumatique, tandis que d'autres restent traumatisés. La persistance de mémoires intrusives, vives et stressantes, est une caractéristique centrale du trouble de stress post-traumatique (TSPT). Cette condition psychiatrique a été longtemps considérée comme un trouble de la mémoire enraciné dans des altérations hippocampiques. Des études récentes ont ainsi proposé que le TSPT soit lié à un dysfonctionnement généralisé du réseau cérébral responsable de la suppression des mémoires intrusives. Le but de cette thèse était d'identifier des marqueurs capables de discriminer et prédire des conséquences résilientes ou pathologiques, suite aux attentats terroristes du 13 novembre 2015 à Paris. Dans une première étude, nous avons utilisé des méthodes de modélisation computationnelle et de connectivité cérébrale pour caractériser deux différents mécanismes, qui sous-tendent le contrôle des mémoires intrusives : le contrôle prédictif et le contrôle réactif. Nous avons trouvé que les personnes qui développaient un TSPT formaient des croyances anormales concernant les mémoires intrusives à venir, accompagnées d'efforts exagérés pour les prévenir, et l'incapacité de se débarrasser des mémoires intrusives qui revenaient à l'esprit. Le déséquilibre entre le control prédictif et le contrôle réactif était corrélé avec une plus grande sévérité des symptômes d'évitement. Dans une deuxième étude, longitudinale, nous avons exploré comment le contrôle de la mémoire et les volumes hippocampiques évoluaient chez les sujets remis du TSPT et les sujets avec un TSPT persistant. Nous avons trouvé que la rémission du TSPT, trois ans après le trauma, était associée à des changements plastiques de l'hippocampe et au rétablissement de l'équilibre entre le contrôle prédictif et le control réactif des mémoires intrusives. Ces deux marqueurs prédisaient ainsi la diminution future des symptômes, cinq ans après le trauma. Ces résultats suggèrent que la plasticité neurocognitive des circuits du contrôle et de la mémoire peut prédire la rémission ou la persistance du TSPT dans le temps.

Keywords: Post-traumatic stress disorder; Memory suppression; Computational psychiatry; Hippocampal subfields; Bayesian modelling; predictive control.