



HAL
open science

Influences du contrôle inhibiteur et des schémas collectifs sur la mémoire individuelle : étude en électrophysiologie et imagerie cérébrale fonctionnelle

Nicolas Legrand

► **To cite this version:**

Nicolas Legrand. Influences du contrôle inhibiteur et des schémas collectifs sur la mémoire individuelle : étude en électrophysiologie et imagerie cérébrale fonctionnelle. Psychologie. Normandie Université, 2018. Français. NNT : 2018NORMC034 . tel-02417046v2

HAL Id: tel-02417046

<https://theses.hal.science/tel-02417046v2>

Submitted on 6 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité PSYCHOLOGIE

Préparée au sein de l'Université de Caen Normandie

Influences du contrôle inhibiteur et des schémas collectifs sur la mémoire individuelle : étude en électrophysiologie et imagerie cérébrale fonctionnelle

**Présentée et soutenue par
Nicolas LEGRAND**

**Thèse soutenue publiquement le 17/12/2018
devant le jury composé de**

M. PHILIPPE FOSSATI	Professeur des universités PraticienHosp, Université Paris 6 Pierre et Marie Curie	Rapporteur du jury
Mme CATHERINE TALLON-BAUDRY	Directeur de recherche, École Normale Supérieure de Paris	Rapporteur du jury
M. PIERRE GAGNEPAIN	Chargé de recherche à l'INSERM, Université Caen Normandie	Membre du jury
Mme MARIA WIMBER	Maître de conférences, Université de Birmingham - Royaume Uni	Membre du jury
M. FRANCIS EUSTACHE	Professeur des universités, Université Caen Normandie	Directeur de thèse
M. DENIS PESCHANSKI	Directeur de recherche au CNRS, Université Paris 1 Panthéon-Sorbonne	Co-directeur de thèse

Thèse dirigée par FRANCIS EUSTACHE et DENIS PESCHANSKI, Neuropsychologie cognitive et neuroanatomie fonctionnelle de la mémoire humaine



UNIVERSITÉ
CAEN
NORMANDIE



Résumé

L'oubli est un mécanisme essentiel au fonctionnement de la mémoire. Ceci est vrai lorsque l'on considère la mémoire de l'individu, qui est l'objet privilégié de la neuropsychologie, mais cela est également le cas lorsque nous considérons les usages sociaux, culturels et historiques de la mémoire qui sont les objets d'études privilégiés de l'histoire et des sciences humaines. Dans ce travail, nous avons cherché à comprendre l'influence conjointe des capacités de contrôle inhibiteur et de l'encodage de représentations collectives sur la conservation des souvenirs épisodiques. Nous avons abordé la question de l'oubli volontaire à l'aide du paradigme "*Think/No-Think*" qui permet l'induction expérimentale d'intrusions mnésiques, de leur contrôle et de leur oubli. Nous avons tout d'abord montré, en électrocardiographie et en électroencéphalographie, que les capacités d'oubli volontaire influencent directement la manifestation physiologique des émotions qui peuvent leur être associées. Ce résultat a ensuite été complété par une étude, à l'aide d'analyses de classification multivariées, montrant que les dynamiques neurales qui encadrent la suppression effective des souvenirs, lorsqu'ils apparaissent involontairement à la conscience, sont comparables aux réflexes de réorientation attentionnelle qui peuvent intervenir lors d'interférences sensorielles ou perceptives. Enfin, nous abordons l'influence des schémas présents en mémoire collective sur l'encodage des souvenirs à travers une étude en imagerie par résonance magnétique fonctionnelle.

Mots-clés: oubli, inhibition, régulation émotionnelle, électrocardiographie, électroencéphalographie, schémas mnésiques, mémoire collective, imagerie par résonance magnétique

Abstract

Forgetting is a central mechanism for memory functioning. This is noticeable regarding individual's memory, which is the privileged topic of neuropsychology, but this is still the case when we consider the social, cultural and historical practices of memory which are the privileged topics of history and social sciences. In this work, we tried to understand the joint influence of inhibitory control capabilities and the encoding of collective representations over episodic memories conservation. We addressed the question of voluntary forgetting through the use of the "*Think/No-Think*" paradigm as a way to induce experimental memory intrusions and their control. First, we report with electrocardiography and electroencephalography that voluntary forgetting capabilities directly affects the physiological roots of associated emotions. This result was then complemented, through the use of multivariate decoding, by a study showing that the neural dynamics surrounding the effective suppression of memories, when they involuntarily intrude into consciousness, are comparable to the reorienting reflex occurring during sensory or perceptive interference. Finally, we assess the influence of collective memory schema on the encoding of memories in a magnetic resonance image study.

Keywords: forgetting, inhibition, emotional regulation, electrocardiography, electroencephalography, memory schemas, collective memory, magnetic resonance imaging

Remerciements

Je remercie tout d'abord les membres du jury qui ont accepté d'évaluer ce travail, Catherine TALLON-BAUDRY et Philippe FOSSATI pour avoir accepté d'en être les rapporteurs et Maria WIMBER pour avoir accepté d'en être examinatrice.

Mes remerciements vont ensuite à Francis EUSTACHE ainsi qu'à Denis PESCHANSKI, tous deux directeurs de cette thèse, pour avoir accepté d'en superviser l'avancée conjointement, en y favorisant et en y incitant cette double influence des sciences humaines et des neurosciences. Au Pr. EUSTACHE, à nouveau, pour m'avoir accueilli dans son équipe à Caen dès 2013 et pour son implication dans l'avancée des différents projets dont les résultats sont présentés ici.

Je remercie ensuite Pierre GAGNEPAIN, qui a été l'encadrant scientifique principal des travaux de ce manuscrit, pour son dévouement à mener à bien ces différents projets et pour l'ensemble des connaissances que j'ai pu acquérir en travaillant à ses côtés. Je remercie également Géraldine RAUCHS, qui m'a intégré aux études sur le sommeil et les rêves lorsque je suis arrivé dans l'unité, et pour sa participation aux études et réflexions que nous avons menées par la suite sur ces thèmes et qui sont présentées en annexe de cette thèse.

Je remercie également les personnes qui, par leur participation à différentes étapes de ce travail, ont permis que celui-ci voit le jour : Olivier ETARD pour avoir apporté la physiologie aux psychologues que nous sommes, Franck DOIDY et Patrice CLOCHON pour la gestion du labo EEG et Florence FRAISSE qui m'a guidé dans les méandres de l'administration lorsqu'il le fallait. Également les quelques 150 personnes qui se sont portées volontaires durant ces quatre années pour participer aux tests que nous leur réservions, parfois long et fatiguants, pour nous avoir prêté leur cerveaux, leur cœur et leur sommeil.

Table des matières

Résumé	iii
Abstract	v
Remerciements	vii
Table des matières	ix
Liste des figures	xi
Avant-propos	xiii
1 Introduction	1
1.1 Ce qui résiste à la mémoire	1
1.2 Un art de l'oubli ?	4
1.2.1 Une taxonomie des oublis	5
1.2.2 Le paradigme "Think/No-Think"	8
2 Première étude	11
2.1 Les émotions et leurs souvenirs	11
2.1.1 Inhibition mnésique et régulation émotionnelle	11
2.1.2 Les bases physiologiques du contrôle des émotions	14
2.2 Étude n°1 - Modulation de l'activité cardiaque induite par la suppression volontaire de souvenirs à valence négative	21
3 Seconde étude	79
3.1 Les dynamiques oscillatoires de la suppression mnésique	79
3.1.1 Intrusions mnésiques et psychopathologies	79
3.1.2 Séquence oscillatoire de l'inhibition mnésique	83
3.1.3 Analyses de patterns multivariées en électroencéphalographie pour la détection d'intrusions	84
3.2 Étude n°2 - Détection multivariée des intrusions mnésiques et étude des processus d'inhibition réactionnels en électroencéphalographie	88
4 Troisième étude	127
4.1 Mémoires individuelles, mémoires collectives	127
4.1.1 Mémoire collective : des neurosciences à l'histoire	127

4.1.2	Trois variantes des mémoires collectives	129
4.1.3	Vers une neuroscience des mémoires collectives	132
4.2	Étude n°3 - Rôle du cortex préfrontal médian dans l'encodage et la récupération de représentations historiques	138
5	Discussion	165
5.1	La permanence du souvenir	165
5.2	Une régulation émotionnelle des souvenirs	167
5.3	Le rythme des intrusions	170
5.4	Oublier l'histoire ?	173
6	Annexes	179
7	Bibliographie	233

Liste des figures

2.1	Variabilité de la fréquence cardiaque	18
3.1	Analyses de patterns multivariées appliquées à l'électroencéphalographie	86
3.2	Généralisation temporelle du décodage multivarié appliquée à différentes bandes de fréquences	87
4.1	Trois notions de mémoire collective	131
4.2	Création et utilisation des schémas mnésiques	136
5.1	Modèle d'intégration neuroviscérale	170
5.2	Dynamiques possibles du contrôle inhibiteur et des intrusions mnésiques dans la régulation des intrusions	172
5.3	Dynamiques possibles du contrôle inhibiteur et des intrusions mnésiques dans la régulation des intrusions	173
5.4	Oubli induit par la reconfiguration des schémas mnésiques	177

Avant-propos

Le travail que nous présentons dans ce manuscrit rassemble quatre études et deux revues de questions qui ont été réalisées durant mon doctorat. Trois études rapportant des résultats expérimentaux sont proposées dans le corps du texte et occupent chacune une partie. Deux articles théoriques rédigés en français et les résultats d'une étude supplémentaire présentée en langue anglaise figurent en annexes et apportent des éclairages complémentaires aux thèmes qui sont abordés dans le manuscrit.

Les deux premiers articles portent sur les capacités de suppression volontaire des souvenirs. Le contrôle inhibiteur est un mécanisme essentiel pour prévenir la réactivation de souvenirs dans des contextes inappropriés. On considère en particulier que cette fonction pourrait jouer un rôle important pour empêcher le développement des symptômes de nombreux troubles psychiatriques et dans la régulation émotionnelle d'une manière générale. Pour étudier ces mécanismes expérimentalement, nous avons utilisé le paradigme "*Think/No-Think*" proposé par Anderson and Green, 2001, qui permet d'induire des intrusions mnésiques et leur contrôle volontaire en laboratoire.

La première étude a cherché à comprendre au moyen d'examen en électroencéphalographie (EEG) et en électrocardiographie (ECG) comment ce processus initié par des aires cérébrales situées dans le cortex préfrontal peut interagir avec le système nerveux autonome pour réguler l'expression d'émotions associées à des souvenirs volontairement supprimés. La seconde étude a abordé les dynamiques oscillatoires de ce mécanisme. En utilisant le décodage multivarié de signaux EEG enregistrés lors de la suppression volontaire de souvenirs, nous avons cherché à isoler la séquence précise d'intrusion et de contrôle des représentations pour évaluer l'implication de différentes bandes de fréquence dans le contrôle effectif de ces intrusions. Les données présentées dans ces études sont issues du protocole ENGRAMME dont j'ai participé à la création sous la supervision du Dr. Pierre Gagnepain. J'ai été impliqué dans les démarches pour les demandes d'autorisation, la conceptualisation des épreuves, j'ai réalisé leur développement informatique, l'acquisition des données dans leur intégralité auprès d'une soixantaine de participants et développé les scripts d'analyse des résultats qui sont présentés dans les articles.

La troisième étude ouvre la thématique de l'oubli au domaine des sciences humaines et sociales et aborde le concept de mémoire collective par le prisme des neurosciences. La mémoire collective désigne un ensemble de représentations partagées par les membres d'un groupe connectés par un passé commun, et qui sont constitutives de leur identité collective. On considère en particulier que ces

représentations sont inscrites dans la mémoire des individus ou dans les ressources culturelles des sociétés sans comprendre exactement comment ces dimensions pourraient interagir. En utilisant une approche transdisciplinaire, nous avons tenté de montrer que les souvenirs collectifs inscrits dans les ressources culturelles d'une société sont également encodés dans certaines aires cérébrales (ici en particulier dans le cortex préfrontal médian) et qu'ils interviennent dans l'encodage et le rappel de nouvelles représentations liées à l'histoire partagée par ces individus. Les données qui sont présentées dans cet article ont été acquises dans le cadre du protocole MULTIBRAIN développé par le Dr. Pierre Gagnepain. J'ai participé à l'analyse des données ainsi qu'aux réflexions théoriques portant sur l'étude de la mémoire collective. La première partie des résultats de ce protocole est par ailleurs en cours de publication mais est indépendante du travail que j'ai fourni et qui est présenté ici.

Ces trois études sont précédées d'une introduction générale et d'une introduction spécifique à chaque article, toutes rédigées en français. Enfin, une discussion générale reprend les principaux résultats portant sur les mécanismes de contrôle de la mémoire, sur un plan individuel comme sur un plan collectif.

” *Il y a un degré d'insomnie, de rumination, de sens historique qui nuit au vivant et qui finit par le détruire, qu'il s'agisse d'un homme, d'un peuple ou d'une civilisation.*

— Nietzsche

(Seconde considération intempestive, §1)

1.1 Ce qui résiste à la mémoire

Du côté de chez Swann est le premier volume d'*À la recherche du temps perdu*, roman écrit par Marcel Proust au début du XX^{ème} siècle (Proust, 1919). Ce texte est notamment célèbre pour la mention d'un épisode qui a marqué, au-delà de la littérature, les rapports que pouvaient entretenir le récit à la première personne avec différentes formes de savoirs scientifiques, et en particulier avec la psychologie et les neurosciences naissantes. Ce fait est à ce point notable qu'aujourd'hui encore, s'il nous fallait trouver dans la littérature l'incarnation la plus parlante des avancées neuroscientifiques autour de la mémoire, l'épisode de la madeleine et les mécanismes phénoménologiques qu'il décrit seraient certainement l'un des points de consensus les plus denses parmi les disciplines concernées. La recherche qu'entreprend le narrateur s'ouvre en effet par la description d'une scène de réminiscence qui se voudrait involontaire, induite par le goût d'une madeleine trempée dans une cuillère de thé, et qui va venir réactiver des souvenirs anciens qui semblaient avoir disparu, au moins temporairement, de la mémoire du narrateur. Ce passage est remarquable par bien des aspects.

Pour la littérature et la phénoménologie d'abord. Cette influence de la matière que représente la madeleine sur les dynamiques intérieures du narrateur n'est pas sans rapports avec le dualisme que proposait Bergson dans *Matière et Mémoire* quelques années auparavant (Bergson, 2012) et a été largement commenté. Mais c'est aussi pour la psychologie naissante que ces circonstances et descriptions fines purent faire écho à ce qui sera qualifié un demi siècle plus tard de souvenir épisodique (Tulving and Thomson, 1973; Tulving, 1972), et dont l'étude reste aujourd'hui encore un vecteur majeur de la recherche fondamentale. Contrairement aux connaissances sémantiques, qui constituent une seconde composante de la mémoire explicite, les souvenirs épisodiques ne sont pas des connaissances abstraites sur le monde dont on ne saurait restituer l'apprentissage originel, mais des expériences vécues à la première personne et qui intègrent un ensemble de modalités sensorielles. Que des scènes remontant à l'enfance du narrateur puissent être réactivées sous la seule impulsion d'une sensation, et notamment s'il s'agit d'un goût ou d'une odeur,

est ainsi une illustration brillante de la profondeur de l'encodage de ces événements, et de leur ancrage associatif et sensoriel.

Cette proximité entre science et la littérature a ainsi encouragé la mise à l'épreuve des théories d'un Proust "neuroscientifique" (Lehrer, 2007) et des "phénomènes proustiens" qui lui étaient associés. Une étude (Herz and Schooler, 2002) a par exemple rapporté, en lien explicite avec le texte de Proust, que les souvenirs autobiographiques évoqués par l'intermédiaire de l'odorat contiennent plus d'émotions que ceux qui sont évoqués par d'autres biais. La plupart des articles scientifiques qui s'attachent à cet exemple emprunté à la littérature (Bray, 2013) l'ont fait en conservant malgré tout une certaine distance avec son interprétation (Troscianko, 2013). Mais, de façon intéressante, ces raccourcis révèlent aussi une certaine représentation de la mémoire.

Un premier exemple pourrait consister à penser que la madeleine illustre la réactivation automatique d'un souvenir émotionnel via l'odorat et le goût. Cette schématisation en deux étapes qui irait directement de l'odeur au souvenir, gomme en effet le rôle et la présence immédiate de l'oubli, ou en tout cas la difficulté à accéder au souvenir, qui intervient en premier lieu dans l'extrait. Le temps dont le narrateur fait de ce récit est un temps qui est perdu. On pourrait considérer que cela est finalement le cas de tout souvenir qui, après avoir été longtemps laissé de côté, est finalement rappelé à la conscience. Mais, la ré-évoquant à laquelle se livre le narrateur est d'abord marquée par l'étrangeté : ce qui se rappelle à lui semble en premier lieu inconnu. Même, ce qui est évoqué par la gorgée de thé n'est certainement pas un souvenir, c'est une émotion, presque une sensation : "Un plaisir délicieux m'avait envahi, isolé, sans la notion de sa cause". Cette simple remarque contraste déjà avec l'a priori selon lequel les odeurs évoquent des souvenirs autobiographiques qui contiennent des émotions plus fortes. Le texte nous pousserait presque à dire l'inverse : que ce sont les émotions qui ont des souvenirs, et qu'ils ne leur sont en fait même pas constitutifs, ni même vraiment bien associés, et encore moins spontanément réactivés.

Un second raccourci pourrait aussi être décelé dans la qualification même de souvenir involontaire. Cette confusion est plus difficile à isoler dans la mesure où l'exploration des richesses des mécanismes involontaires de l'esprit tient quasiment lieu de méthode dans l'ensemble du roman. (Schacter et al., 1998) parlent par exemple des expériences de réminiscence du roman comme des actes involontaires et non intentionnels de récupération "ne demandant pas d'efforts délibérés de se projeter dans le passé". Cette affirmation est pourtant assez contestable lorsque l'on réalise que l'essentiel des descriptions qui suivent sont celles d'un travail mental acharné pour retrouver les causes d'une émotion spontanément (ré)apparue sous l'influence d'une odeur et d'un goût. Proust ici parle en effet d'un temps qui a été retrouvé, et ce résultat est le fruit d'un effort. La majeure partie de cet épisode qui devrait illustrer l'automatisation d'un processus de rappel est en réalité consacrée à la description des tentatives longtemps infructueuses du narrateur pour retrouver l'origine d'une émotion : "quelque chose qu'on aurait désancré, à une grand profondeur". Il y a donc

deux oublis : un premier qui serait dû à l'affaiblissement des traces mnésiques au cours du temps, que l'on pourrait qualifier d'involontaire. Et un second, qui menace potentiellement le souvenir et qui pourrait être qualifié de volontaire. C'est au narrateur de décider s'il rappelle ou non cette impression vague à sa mémoire : "*Qui sait s'il remontera jamais de sa nuit ?*". On voit donc assez peu d'instantanéité et de facilité dans la réminiscence. Il faut au contraire diverses tentatives pour partir à la recherche de ces souvenirs, les attiser, les préparer pour le rappel. Et si ce choix-là est possible, il faut reconnaître que le contraire l'est tout autant. Pourquoi ne pas alors résister aux souvenirs, les repousser, et taire leur ressemblance avec l'instant présent ? Peut-on même les renvoyer dans leurs profondeurs s'ils sont déjà là ? Proust nous parle du rôle central des émotions dans le processus de rappel, mais aussi du pouvoir considérable de la volonté sur les éléments de la mémoire : pouvoir de rappel et pouvoir d'occultation.

Le contrôle inhibiteur est un processus cognitif engagé dans la diminution ou la suppression complète d'autres processus cognitifs, en particulier lorsque ceux-ci sont initiés de façon automatique et qu'ils peuvent interférer avec les objectifs ou les comportements qui sont visés par l'individu. Le contrôle inhibiteur est un élément central lorsqu'il s'agit d'empêcher un souvenir de réapparaître à la conscience ou de réduire l'expression d'un état émotionnel inapproprié par rapport à un contexte défini. De tels mécanismes sont pour ces raisons fortement impliqués dans le maintien de la santé mentale d'un individu, mais sont aussi étudiés depuis une quinzaine d'années pour leur rôle dans la suppression volontaire de souvenirs indésirables, et en particulier s'ils véhiculent des émotions fortement négatives. Le travail que nous présentons ici porte sur l'oubli induit par l'inhibition du rappel. Cet oubli est pris comme une mécanique inhérente à la mémoire elle-même, ce qui le distingue d'autres types d'oublis que nous allons détailler dans la partie suivante. Cette forme d'oubli est aussi ce qui encadre l'épisode même de la madeleine de Proust : celui des souvenirs longtemps occultés qui nous paraissent étrangers, et celui du choix, volontaire, de laisser dans la profondeur, peut-être pour toujours, ces éléments qui sont évoqués mais non encore ressurgis. Nous nous intéresserons en particulier à ce phénomène d'intrusion mnésique, le rappel involontaire de souvenirs épisodiques en réponse à un élément déclencheur, dans l'environnement ou dans les pensées mêmes du sujet. Si de telles intrusions peuvent avoir des aspects plaisants comme dans le passage que nous avons évoqué, la difficulté à les contrôler lorsqu'elles véhiculent des émotions négatives est un trait partagé par de nombreuses psychopathologies.

Dans les études 1 et 2, nous avons utilisé le paradigme "*Think/No-Think*" (Anderson and Green, 2001) pour reproduire expérimentalement l'intrusion involontaire de souvenirs. La **première étude** porte sur l'oubli, les émotions et leurs souvenirs. Nous y rapportons à travers deux expériences que la suppression volontaire de souvenirs émotionnels influence non seulement les représentations mnésiques, mais également l'évaluation émotionnelle et la réaction physiologique associées à ces émotions, ce qui nous pousse à proposer un modèle de contrôle cognitif intégrant bien plus largement les deux branches du système nerveux : central et autonome.

La **seconde étude** s'intéresse plus particulièrement aux dynamiques opposant le contrôle volontaire des traces mnésiques à leurs intrusions involontaires dans la conscience d'un individu. Nous y rapportons les résultats d'une étude en électroencéphalographie (EEG) réalisée lors de la résistance volontaire à la ré-évoation de souvenirs contenant des émotions négatives. En particulier, nous nous intéressons à la séquence cognitive et aux mécanismes qui sous-tendent le rejet effectif de ces intrusions mentales. Enfin, la **troisième étude** ouvre ces problématiques à un champ plus large : celui de l'occultation et de la conservation des souvenirs qui se rattachent à une identité historique et qui intègrent pleinement les cadres sociaux dont parlait le sociologue français Maurice Halbwachs (Halbwachs, 1925), sous le regard croisé des neurosciences et des sciences sociales. L'idée selon laquelle les structures sociales dans lesquelles nous sommes intégrés peuvent influencer l'organisation de la mémoire, y compris la mémoire épisodique des événements directement vécus, est présente de longue date dans la psychologie. Le rappel auquel aboutit l'épisode de la madeleine illustre cette réapparition progressive d'éléments qui entretiennent entre eux des liens schématiques qui vont structurer ce qui sera un récit progressif et non un tableau. Ce que le narrateur évoque alors n'est pas un souvenir isolé, mais un ensemble de souvenirs pris dans une structure qui a intégré entre temps des éléments sociaux et contextuels variés. Nous présenterons alors les résultats d'une étude en imagerie par résonance magnétique fonctionnelle portant sur les bases cérébrales impliquées dans l'encodage et le rappel d'événements historiques appartenant à la mémoire collective. Nous détaillerons un modèle neurobiologique de cette influence culturelle, historique et sociale de la mémoire et proposerons, en discussion, l'intégration de ces perspectives avec l'étude des mécanismes de l'oubli précédemment évoqués.

1.2 Un art de l'oubli ?

Oubli et mémoire ont souvent été placés aux deux extrémités d'un axe représentant la perfection et les échecs de la conservation des informations. Les arts de la mémoire (*ars memoriae*) étaient alors toute technique capable de rendre cette conservation moins faillible. On conçoit alors difficilement l'idée qu'il puisse y avoir un art de l'oubli (*ars oblivionalis*), parce qu'on en verrait pas l'utilité, mais aussi parce qu'on en voit mal la possibilité. Umberto Eco dans un article célèbre sur la question (Eco and Migiel, 1988), rejetait entièrement la possibilité conceptuelle d'une telle pratique et la classait volontiers au côté d'autres disciplines impossibles. Pourtant, cette vision est loin d'avoir dominé l'histoire de la psychologie, qui a rapidement vu dans l'oubli à la fois le processus même de la construction mnésique (on retient ce qui n'a pas été oublié), et également une forme de protection psychique (on vit mieux en oubliant). Le signal a d'ailleurs été donné par l'un des pères fondateurs de la discipline en personne, le psychologue américain William James, lorsqu'il affirmait dans ses cours de psychologie que "*dans l'utilisation pratique de notre intellect, oublier est aussi important que se souvenir*" (James, 1892). On trouvera des considérations similaires, quoiqu'ayant une portée différente, dans la *Seconde*

considération intempestive de (Nietzsche, 1874), dont les conceptions auront aussi une influence considérable sur ce que Freud pourra dire ensuite du refoulement et de l'inconscient. L'oubli pris comme faculté a donc une histoire assez ancienne. Pour autant, proposer une telle alternative à l'axe unidimensionnel mémoire-oubli implique un certain nombre d'avancées déterminantes tant pour la psychologie que pour la physiologie concernant notamment les différents systèmes de mémoire et le rôle du système nerveux central dans l'encodage des souvenirs. Il importera donc de pouvoir dissocier différentes formes d'oubli, nous en évoquerons trois, et de pouvoir les renvoyer à un fonctionnement de la mémoire bien défini.

1.2.1 Une taxonomie des oublis

S'il fallait en fournir une définition provisoire, nous pourrions dire de l'oubli dans un premier temps ce qu'il affecte : les traces, sous l'une des formes qu'évoque (Ricoeur, 2003) dans *La mémoire, l'histoire, l'oubli* : neurale, mnésique (consciente ou inconsciente) ou encore écrite. Et plus précisément encore que ce mécanisme consiste à mettre ces traces hors de portée d'un système - neural, cognitif ou sociétal - qui était en mesure auparavant de lui donner du sens. La destruction pure et simple des traces vaut évidemment oubli. Mais on remarque immédiatement que des formes dérivées peuvent être tout aussi radicales. Occulter les traces, au sens où elles ne seraient plus accessibles au système interprétant, vaut également oubli. Mais toute évolution aussi, qui viendrait à changer l'interprétation même que fait le système vivant des traces, vaudrait oubli, c'est en ce sens que Nietzsche dira que l'oubli est nécessaire à la vie même. Le remplacement des traces, l'orientation de l'attention sur autre chose, le divertissement de l'homme dont nous parle Pascal, sont des exemples où la qualification d'oubli n'est pas qu'une formule. Si l'oubli n'est pas que l'opposé de la mémoire, il semble alors en exister une large variété. Mais, s'il existe, nous considérons aussi qu'il est difficile à mettre en œuvre, et qu'il ne se réalise en définitive que très rarement. Les expériences négatives, voire des traumatismes, montrent que cette idée d'un oubli volontaire n'est pas toujours évidente. Nous allons ici dissocier trois formes de mise en œuvre de l'oubli, telles qu'elles ont pu être observées par la psychologie et les neurosciences, en tentant de respecter une gradation dans l'implication d'une certaine forme de volonté : nous parlerons d'un oubli par déclin, d'un oubli par sélection, et enfin d'un oubli induit qui sera la forme qui nous concernera pour la suite.

La première forme est certainement la plus intuitive, il s'agit de l'**oubli par déclin** de la trace mnésique. C'est de cet oubli dont parle le psychologue allemand Hermann Ebbinghaus lorsqu'il propose entre 1880 et 1885 qu'il est possible de décrire via une "courbe de l'oubli" la disparition progressive de sa mémoire de syllabes qu'il s'était fait lui-même apprendre (Ebbinghaus, 2011). L'oubli renvoie alors à un mécanisme involontaire d'affaiblissement progressif des traces mnésiques : le destin de toute information encodée en mémoire est ainsi de complètement disparaître un jour, à moins que celle-ci ne soit consolidée à intervalles réguliers. Cette vision sera largement étayée par les avancées de la neurobiologie au cours du XX^{ème} siècle, et sera reprise jusque dans les sciences humaines et sociales puisque

de telles courbes de l'oubli font partie de certains concepts qui ont aisément pu être transférés à l'étude de la mémoire collective (Roediger and DeSoto, 2014).

La seconde forme que nous pouvons évoquer est celle d'un **oubli par sélection** de la trace mnésique. Cette notion fut fondatrice pour la définition même de l'acte de mémoire puisqu'elle souligna que la conservation d'un souvenir se fait autant par un encodage de la trace mnésique que par la suppression d'autres éléments interférants. William James, dans la référence citée plus haut, renvoie à cette nécessité de la disparition d'une majorité des éléments de l'expérience pour permettre la conservation de souvenirs compréhensibles pour le sujet. C'est ce dont traite Borges dans *Funès ou la mémoire* en décrivant la démesure induite par l'hypermnésie de son personnage et les difficultés qu'il rencontre même pour penser, abstraire, occulter les détails de certains événements pour en garder une structure générale (Borges, 1993). L'oubli induit par la sélection et le tri des traces mnésiques fait également partie des paradigmes les plus repris tant par les neurosciences que par les sciences humaines et sociales. La littérature qui a accompagné depuis le milieu des années 1990 la mise en évidence du rôle du sommeil dans la consolidation des traces mnésiques acquises durant la veille a également souligné son rôle déterminant dans la suppression de souvenirs dispensables (Stickgold and Walker, 2013).

Ces deux formes d'oubli renvoient jusqu'à présent à des mécanismes relativement automatisés, on pourrait même dire que le temps est le facteur principal et que le comportement du sujet paraît secondaire. Nous pouvons alors évoquer une troisième forme de l'oubli, qui serait dépendante des événements extérieurs au sujet, mais aussi et surtout influencée par ses états mentaux et susceptible d'être orientée d'une certaine manière par la force de la volonté. De nombreuses expériences en psychologie ont pu montrer ce type de mécanisme. Les réunir sous un seul concept n'est pas chose aisée tant les processus sur lesquels ils reposent peuvent être variés. Ainsi parler d'oubli "volontaire", "intentionnel" ou encore "dirigé" est sans doute trop spécifique. Nous préférons alors suivre l'expression anglaise d'**oubli induit** ("*induced forgetting*"). Ce terme permet de comprendre qu'à la différence de l'oubli par déclin, qui est un processus passif d'oblitération progressif de la trace mnésique, ou encore de l'oubli par sélection, qui est partie intégrante du processus même de formation des souvenirs, un oubli induit désignera la suppression d'un souvenir en réaction à un état extérieur ou intérieur du sujet. Ces mécanismes opèrent également un déplacement sur le plan neurobiologique puisqu'il n'est plus ici question de la mécanique physiologique d'encodage ou de sélection, mais de l'influence que le système cognitif plus largement peut exercer sur les souvenirs encodés essentiellement via les fonctions exécutives, et notamment l'inhibition. Le premier exemple d'un tel oubli que nous pouvons mentionner est l'**oubli induit par l'inhibition de l'encodage** encore appelé "*oubli dirigé*". La dimension volontaire, sélective et intentionnelle de cet oubli décrit pour la première fois par (MacLeod, 1975) en a fait un mécanisme largement exploré par la psychologie cognitive et les neurosciences. Ce paradigme repose sur la présentation d'items au participant, suivie d'une instruction lui demandant de se rappeler ou d'oublier ce qui vient de lui

être présenté, l'instruction pouvant apparaître après chaque item ou après une liste d'items selon la version de la tâche. De façon déterminante, les items pour lesquels il est demandé au participant d'oublier sont effectivement moins bien retenus à la fin de l'épreuve. Il est à noter ici que cet oubli repose sur une inhibition de l'encodage : les éléments qu'il faut oublier ne sont pas encore stockés sous leur forme définitive en mémoire à long terme, et l'idée selon laquelle ce mécanisme pourrait s'appliquer par exemple à la suppression de souvenirs traumatiques dans le cadre de troubles définis doit de ce fait se faire avec précaution. Cette épreuve a en revanche l'avantage de mettre en évidence l'influence des fonctions exécutives dans le mécanisme d'encodage et révèle un phénomène de tri qui est modulé par certains états mentaux et la volonté du participant à conserver ou à oublier le souvenir.

Une seconde modalité d'oubli induit que nous pouvons mentionner est l'**oubli induit par la récupération** mis en évidence par Anderson et al., 1994. Ce paradigme a permis de mettre en évidence le fait que tout rappel n'est pas un acte neutre, mais implique l'occultation de souvenirs liés mais que l'on ne souhaite pas alors rappeler. Après que le sujet ait appris un ensemble d'associations, par exemple entre des mots et des images appartenant à des catégories différentes, il est demandé au participant de se souvenir de certains éléments alors que d'autres sont laissés de côté. De façon déterminante, lors d'un rappel global qui a lieu par la suite, on remarque que la réactivation d'un item dans une catégorie a induit une moindre réactivation des items associés à la même catégorie. Plus précisément, il a fallu la mise en œuvre d'un processus d'inhibition pour permettre la récupération de l'information sans interférence avec d'autres éléments qui étaient proches. Cette inhibition ne sera en revanche pas nécessaire pour les items appartenant à d'autres catégories et qui n'interféreront pas avec le rappel. Cette expérience permet donc de mettre en évidence un oubli paradoxal puisqu'il concerne les items appartenant aux catégories les plus rappelées. Ce phénomène a été repris et adapté dans le cadre d'études portant sur la dimension sociale et partagée de la construction des souvenirs. Le rappel répété de certains éléments d'un événement est un trait central dans le traitement médiatique ou les discussions informelles autour d'éléments de l'actualité. En appliquant ce même mécanisme à la ré-évocation de souvenirs lors de discussions, (Cuc et al., 2007) ont par exemple mis en évidence l'existence d'un **oubli induit par la récupération partagé socialement** ("*Socially-Shared Retrieval Induced Forgetting*"), montrant une influence sur la mémoire de celui qui prend la parole autant que sur celle de ceux qui écoutent. Nous détaillerons davantage les travaux qui ont suivi le développement de cette notion dans la Partie 3 de ce travail.

Enfin, il existe une troisième forme d'oubli induit, et elle concernera les deux premières parties de ce travail. Il s'agit de l'**oubli induit par l'inhibition du rappel**, qui fut décrit pour la première fois par (Anderson and Green, 2001), en même temps que l'expérience qui en avait permis l'observation : le paradigme "Think/No-Think".

1.2.2 Le paradigme "Think/No-Think"

A la différence des phénomènes évoqués précédemment, le paradigme "Think/No-Think" a mis en avant la capacité de suppression volontaire de souvenirs encodés en mémoire à long terme via l'inhibition du rappel de ceux-ci lorsqu'ils sont évoqués par des indices associés. La qualification même de suppression volontaire rapproche quasi immédiatement cette modalité du concept freudien de "refoulement", et c'est d'ailleurs dans un contexte de controverses intenses avec la psychanalyse que ce mécanisme fut révélé par la psychologie expérimentale. De sa mise à l'écart du processus de fondation des sciences cognitives initié par la réunion célèbre le 11 septembre 1956 au MIT, jusqu'à sa réintégration dans les départements de psychologie dans les années 1980, l'héritage freudien s'est longtemps construit dans un rapport alternant adhésion et conflits envers la rationalité des sciences du cerveau émergentes (Castel, 2006). L'inconscient, et son corollaire le refoulement, étant deux des concepts les plus débattus alors. L'année 1995 fut marquée par la publication d'un livre, *The Memory Wars: Freud's Legacy in Dispute* (Crews, 1995), qui donna à ces controverses une tournure médiatique singulière. Les années précédentes avaient été marquées aux États-Unis par un ensemble d'attaques envers les thérapies centrées sur la récupération de souvenirs qui avaient alimenté la littérature avec un ensemble de cas d'abus sexuels considérés comme refoulés et qui avaient ressurgi lors de séances d'hypnose à l'âge adulte. Cinq ans auparavant, la psychologue américaine Elizabeth F. Loftus était intervenue lors du procès de George Franklin, accusé par sa fille de l'assassinat d'une de ses amies lorsqu'elle avait 9 ans. La fiabilité de souvenirs retrouvés à la suite de thérapies était au centre du débat et, se basant sur ses propres recherches, Loftus invalida le témoignage principal et mit en avant la malléabilité de la mémoire et la création de faux souvenirs.

C'est dans ce contexte, où la fiabilité des témoignages autour d'événements considérés comme refoulés et obtenus lors de thérapies était intensément questionnée, que (Anderson and Green, 2001) ont rapporté dans la revue *Nature* la démonstration expérimentale d'un oubli définitif mis en œuvre par le participant lui-même. Ces auteurs ont montré pour la première fois que le fait d'éviter consciemment la réactivation d'un souvenir, de le repousser hors de la conscience, de le laisser comme dirait Proust dans sa profondeur a tendance à en induire insidieusement la suppression. Après avoir appris un ensemble de paires de mots, il était demandé au participant de rappeler certaines associations lorsque le mot indice était présenté en vert, ou de tenter par tous les moyens de ne pas le récupérer en mémoire si le mot était présenté en rouge. Cette protection du système cognitif contre la réactivation d'un souvenir ciblé est un mécanisme central de la régulation émotionnelle, et un facteur de défaillance commun à de nombreux troubles psychiatriques. Mais de façon déterminante, les participants étaient non seulement de plus en plus performants pour contrôler leurs souvenirs au fur et à mesure qu'ils progressaient dans la tâche, mais il fut rapporté que lors d'un rappel global réalisé à la fin de l'épreuve, les mots associés aux indices apparus en rouge semblaient avoir effectivement disparu de leur mémoire. Ce résultat fut connecté, par les auteurs eux-mêmes, au concept freudien de refoulement.

Cette notion est encore largement débattue par les neurosciences (Naccache, 2006; Panksepp and Solms, 2012), et la question de la permanence des souvenirs après les processus de suppression est trop étendue pour qu'elle soit résumée ici. Nous nous contenterons de mentionner la postérité qu'a connu ce paradigme. La revue proposée par (Anderson et al., 2011) faisait alors état de 32 répliques du résultat principal (i.e. une suppression des souvenirs s'il est demandé au participant d'inhiber le rappel de ceux-ci) sur un total de 3299 participants jeunes et ne présentant pas de troubles psychiatriques. Les études en imagerie par résonance magnétique fonctionnelle ont révélé l'implication d'une zone spécifique, le cortex préfrontal dorsolatéral droit, mobilisé dans l'inhibition d'une autre zone du cerveau, l'hippocampe, impliqué dans la récupération des souvenirs épisodiques (Anderson, 2004; Depue et al., 2007). Enfin, l'analyse de l'influence réciproque de ces zones dans le temps a révélé que cette inhibition du rappel a également une influence sur les zones où sont stockés les souvenirs, par exemple le cortex visuel pour les images, et que l'oubli qui est observé ici semblerait correspondre à une suppression volontaire et définitive du contenu des représentations en mémoire qui sont ciblées.

2.1 Les émotions et leurs souvenirs

2.1.1 Inhibition mnésique et régulation émotionnelle

Les émotions font partie d'un héritage évolutif déterminant si nous cherchons à comprendre tant les capacités d'adaptation humaines à des environnements changeants et potentiellement hostiles que le développement et l'expression de troubles psychiatriques pouvant prendre des formes très variées (Panksepp, 1998). Au premier abord, les émotions peuvent être perçues comme un ensemble de réactions sur lesquelles la volonté et des formes de raisonnement élaborées ont peu de prises. Elles constituent ainsi davantage des états induits plutôt que des états choisis par celui qui les éprouve. Une description plus analytique distinguera différentes composantes dans la séquence de réactions caractérisant l'émotion, comprenant des réactions motrices, physiologiques, cognitives et phénoménologiques. Les comportements et les réactions motrices qui résultent par exemple d'un état de peur ou de surprise font partie de l'essence même de ces états (*motio* en latin désignant l'action de se mouvoir). Une émotion est une réaction de l'organisme face à une situation de menace ou de récompense orientée vers sa protection et opérant de façon automatisée, involontaire et parfois inconsciente.

Ces états sont des éléments fondamentaux de la survie de tout organisme, et ont conservé une grande prégnance dans la cognition humaine. Pourtant, les avantages évolutifs que fournissent les émotions ont également un coût. Ce coût est celui du contrôle et de la modération, absolument nécessaires, de comportements qui, s'ils interviennent en décalage avec le contexte, peuvent s'avérer au moins contre-productifs, sinon coûteux pour l'organisme ou même directement représenter un danger pour la survie de l'individu. Les troubles anxieux montrent par exemple une telle mobilisation de mécanismes de défense en décalage avec la menace objective du contexte. On parlera alors de *régulation émotionnelle* comme la mise en place de processus permettant de réduire l'impact des émotions négatives (missing citation) sur les comportements mais également leurs corrélats physiologiques, cognitifs et phénoménologiques. L'inadéquation de l'émotion à un contexte donné peut avoir plusieurs origines. (Etkin et al., 2015) distinguent par exemple la perception, l'évaluation et la mise en place d'une réaction dans la construction d'une réaction émotionnelle. Une émotion peut ainsi être indûment initiée par une mauvaise détection ou par la mauvaise interprétation d'un indice de l'environnement. Une émotion peut aussi être justement initiée, mais sans être exprimée de façon cohérente à travers les comportements. Différents mécanismes de dysfonctionnement et de régulations peuvent intervenir lors de ces étapes.

Les déficits dans cette capacité de régulation émotionnelle sont ainsi des éléments centraux dans le développement de nombreuses psychopathologies (Gross and

Jazaieri, 2014; Aldao et al., 2010). Pourtant, les neurosciences cognitives ont souvent envisagé la régulation émotionnelle comme l'ensemble des mécanismes permettant la poursuite d'un comportement orienté vers un objectif en dépit de l'induction d'états émotionnels d'**origine exogène** (Gross, 2015). On considérera alors la capacité d'un individu à contrôler la peur générée immédiatement par la présentation d'une image d'insecte ou de serpent et à pouvoir poursuivre une tâche dans laquelle il était engagé. Mais les émotions négatives qui interfèrent avec la cognition peuvent avoir d'autres origines et être en partie dissociées des éléments directement présents dans l'environnement. La réaction d'un individu à ce stimulus peut être modulée, soit par une habituation, soit par un trauma lié à cette situation. Les souvenirs qui encodent l'histoire de l'individu sont alors des composantes importantes de la construction de l'émotion (Daselaar et al., 2007), et ont au moins une grande influence, sinon la plus grande influence, dans la génération des états émotionnels qui peuvent se déclencher dans certaines pathologies. L'exemple de la madeleine de Proust illustre cette prépondérance de l'affectif dans la ré-évoation : l'émotion la plus importante dans ce passage ne provient pas directement d'un élément de l'environnement mais d'un souvenir. Les ruminations propres aux troubles dépressifs, ou les épisodes de réminiscence lors de stress post-traumatique sont des exemples d'états émotionnels induits en partie au moins par une composante **endogène**, soulignant l'importance de l'histoire de l'individu relativement au stimulus dans la construction de l'état émotionnel qui lui est associé. De façon analogue, la simulation mentale d'événements futurs, qui se base largement sur des mécanismes très proches de ceux du rappel (Schacter et al., 2007; Schacter and Addis, 2007), est une composante centrale dans la production des craintes propres à l'état d'anxiété.

Il a ainsi été récemment proposé que les études neuroscientifiques sur la régulation émotionnelle pourraient largement bénéficier d'une considération plus importante pour les émotions induites par des souvenirs et par leurs intrusions involontaires (Engen and Anderson, 2018). En particulier, la suppression volontaire et le contrôle inhibiteur exercés sur les souvenirs, tels qu'ils sont abordés par la paradigme "*Think/No-Think*", pourraient exercer un rôle dans la régulation émotionnelle encore largement sous-estimé. Les études en imagerie dans ce domaine se sont pour l'instant davantage concentrées sur les stratégies de réévaluation des émotions plutôt que sur toutes les autres stratégies mises ensemble (Morawetz et al., 2017). L'interaction entre la valence des souvenirs et la facilité avec laquelle ceux-ci peuvent être contrôlés volontairement a toutefois été abordée dans le cadre de recherches sur la suppression volontaire. D'après notre revue de la littérature, le paradigme "*Think/No-Think*" a été utilisé au sein de 14 études qui ont intégré un matériel comportant des stimuli de valences différentes, que ce soit via des approches purement comportementales (Lambert et al., 2010; Murray et al., 2011; Hertel and Gerstle, 2003; Joormann et al., 2009; Hertel and McDaniel, 2010; Herbert and Sütterlin, 2012; Hulbert and Anderson, 2018; Depue et al., 2006; Joormann et al., 2005; Nørby et al., 2010), des études en imagerie fonctionnelle (Gagnepain et al., 2017; Liu et al., 2016; Depue et al., 2007) ou encore en imagerie spectroscopique

proche infrarouge (Dieler et al., 2010). Aucune étude n'a jusqu'à présent combiné, ni l'électroencéphalographie, ni la magnétoencéphalographie à ce type d'approche. Les études 1 et 2 du présent travail sont donc les premières à rapporter ce type de données. La tendance générale de ces résultats reste quelque peu contrastée. Certaines études ont montré, de façon sans doute contre-intuitive, un plus grand effet de suppression ou de contrôle sur les souvenirs à contenu émotionnel négatif comparé à des items neutres ou positifs (Depue et al., 2006; Lambert et al., 2010; Joormann et al., 2005; Gagnepain et al., 2017). D'autres études ont rapporté une suppression équivalente (Hulbert and Anderson, 2018; Murray et al., 2011) ou une suppression moindre (Hertel and Gerstle, 2003; Nørby et al., 2010). Dieler et al., 2010 n'ont pas montré de différence comportementale mais une plus grande activité liée au dlPFC droit pour les items émotionnels, suggérant cependant un plus grand recrutement inhibiteur. La variabilité observée s'explique assez largement par le type de matériel utilisé, la valence émotionnelle des items, la modalité d'examen ou encore la population d'étude. La suppression facilitée des souvenirs négatifs apparaît toutefois comme un résultat consistant et majoritairement rapporté.

Ces résultats semblent contredire d'autres observations montrant l'effet important des émotions sur la profondeur de l'encodage des traces mnésiques et leur plus grande conservation dans le temps (Yonelinas and Ritchey, 2015), ce qui devrait impliquer que ces souvenirs soient également plus difficiles à oublier. Il faut pourtant noter que les approches qui mettent en évidence ces résultats n'impliquent pas le contrôle effectif des souvenirs et manipulent *in fine* la qualité de l'encodage de celui-ci. Il est par exemple difficile de faire la part entre l'émotion associée au souvenir et la plus grande attention qu'aura porté le participant à des stimuli de type émotionnel (positifs ou négatifs) plutôt qu'à des stimuli neutres. Si l'on considère néanmoins que ce degré d'attention est resté équivalent, et que les souvenirs ne diffèrent que par les émotions auxquelles ils sont associés, il apparaît pourtant cohérent que le contrôle inhibiteur mobilisé lors de la suppression volontaire d'une intrusion dans la conscience soit davantage biaisé envers le contenu émotionnel des souvenirs pour permettre leur régulation. Le cortex préfrontal dorsolatéral droit, qui est une structure clé dans la mise en œuvre de l'inhibition mnésique (Gagnepain et al., 2017; Gagnepain et al., 2014; Depue et al., 2007) est également une partie centrale dans la régulation émotionnelle (Depue et al., 2015). Il est notamment apparu que le contrôle de souvenirs négatifs mettait en jeu une régulation parallèle de zones traditionnellement impliquées dans l'encodage des souvenirs comme l'hippocampe (Benoit et al., 2015) et de zones impliquées dans l'encodage des émotions qui y sont associées, à savoir l'amygdale (Gagnepain et al., 2017). Ce type de contrôle exercé sur les souvenirs à contenu émotionnel doit être néanmoins distingué d'un autre mécanisme avec lequel il peut être confondu, à savoir le contrôle de l'expression des émotions. Les tentatives de contrôle de l'expression des émotions sont associées à des effets parfois paradoxaux sur la composante subjective des émotions (Goldin et al., 2008), et l'utilisation de cette stratégie se trouve plus souvent liée à des difficultés de régulation émotionnelle (John and Gross, 2004). Les

études basées sur le paradigme "*Think/No-Think*" suggèrent à l'inverse qu'inhibition mnésique et inhibition émotionnelle engagent des mécanismes qui sont largement partagés et qui peuvent agir en parallèle sur les structures encodant les souvenirs et leurs émotions (Gagnepain et al., 2017).

Pourtant, alors même que l'amygdale est une zone sous-corticale dont le rôle est largement établi dans l'encodage de la valence des souvenirs, la question de la suppression "effective" des émotions qui y sont associées paraît ne pas pouvoir être complètement évacuée. En particulier, nous avons mentionné la fonction principalement motrice, de mobilisation des ressources de l'organisme, dans la construction des états émotionnels. Les réponses physiologiques qui sont associées aux émotions paraissent avoir un ancrage à la fois plus important, car initiant la construction cognitive et phénoménologique qui en résultera, et plus résistant car intervenant à un niveau du système nerveux, le système nerveux autonome, dont l'indépendance vis-à-vis des opérations volontaires de l'esprit paraît comme indépassable.

2.1.2 Les bases physiologiques du contrôle des émotions

Les émotions permettent une adaptation rapide à des événements qui pourraient menacer la survie de l'individu. Ces changements se manifestent par des réactions motrices qui nécessitent une adaptation des rythmes physiologiques afin que l'organisme dispose des ressources suffisantes pour initier une défense ou une fuite potentielle. L'accélération du rythme cardiaque et de la fréquence respiratoire sont des phénomènes inhérents à ces changements. L'augmentation du niveau de sudation, la contraction de l'iris ou le ralentissement du système digestif font également partie, quoique moins manifestes, d'une adaptation au niveau de menace que représente l'environnement. Mais ces états ont aussi en retour une influence sur la construction subjective et les processus cognitifs qui sont associés aux émotions. Dans un article célèbre, le psychologue William (James, 1884) a ainsi défini l'émotion comme la perception et les sentiments associés aux changements corporels induits par un stimulus excitateur. Le rôle des signaux afférents, provenant du système nerveux périphérique et remontant vers le système nerveux central, a aujourd'hui été pleinement intégré dans l'étude neuroscientifique des émotions comme, par exemple, dans la théorie des marqueurs somatiques proposée par (Damasio, 1999). On ne saurait en effet réduire les mécanismes émotionnels à la seule activité de l'amygdale sans prendre en compte la quantité d'informations en provenance de divers organes du corps qui est intégrée dans cette structure avant d'être traitée par d'autres zones corticales. Les émotions résultent en grande partie de processus de régulation, de synthèse et de contrôle d'informations sur l'état de l'organisme. C'est le cas par exemple de nombreux signaux en provenance du système cardiovasculaire qui peuvent directement influencer le traitement du sentiment de peur ou d'anxiété (Garfinkel and Critchley, 2016; Critchley and Garfinkel, 2018) et venir biaiser les processus cognitifs liés à ces émotions. A chacun de ses battements, le cœur fait notamment remonter un influx qui est traité par des aires corticales impliquées dans la conscience de soi, l'attention ou le traitement émotionnel, et

dont l'influence peut être étudiée par l'analyse de l'activité électrique ou magnétique du cerveau juste après les battements (*heartbeat evoked potential*, HEP) (Luft and Bhattacharya, 2015; Babo-Rebelo et al., 2016b; Sel et al., 2016; Babo-Rebelo et al., 2016a). Mais cette composante afférente de la construction des états émotionnels ne constitue pas le seul champ d'étude de la psychophysiologie. Un second domaine d'étude également extrêmement actif concerne les mécanismes contradictoires à cette intégration corticale des signaux afférents, à savoir la modulation - par des aires cérébrales potentiellement élevées dans la hiérarchie corticale - de l'activité du système nerveux autonome. Ainsi que l'avait proposé James, une perception qui ne serait pas accompagnée de changements corporels ne nous apparaîtrait que comme un état cognitif abstrait, blafard et dénué de toute affectivité. Poursuivant ce raisonnement, on peut ainsi considérer que toute régulation émotionnelle effective devrait également être une régulation de l'activité physiologique induite par la perception de l'événement et qui en sous-tend la construction cognitive. Cet argument implique pourtant la révision d'une distinction théorique fondamentale entre système nerveux central et système nerveux autonome : la régulation physiologique des émotions implique que ce dernier système n'est pas coupé de l'influence de mécanismes de contrôle volontaire, et par conséquent n'est pas entièrement autonome. Les mécanismes qui permettraient cette régulation physiologique commencent aujourd'hui à être mieux cernés, notamment en ce qui concerne l'une des voies majeures de l'expression des émotions : l'activité du muscle cardiaque.

Le nerf vague

Le système nerveux humain comporte deux grandes composantes : l'une qualifiée de **centrale** qui comprend l'encéphale et la moelle épinière, et une seconde - le système nerveux **périphérique** - intégrant les nerfs et ganglions situés à l'extérieur et ne disposant pas de la protection osseuse de la boîte crânienne ou de la colonne. Ce dernier système peut à nouveau être divisé en **système nerveux somatique** - impliqué dans les traitements sensoriels et moteurs - et **système nerveux autonome**, ou végétatif, dont la fonction est de contrôler les organes internes tels que le cœur, les poumons ou les intestins. L'activité du système nerveux autonome est un indicateur robuste des états émotionnels (Kreibig, 2010). On considère en particulier que deux de ses sous-composantes exercent des influences opposables : le système nerveux **sympathique** d'une part, qui est responsable des réactions de fuite ou de combat et qui va globalement augmenter la capacité de réponse comportementale de l'organisme par une augmentation de son activité (rythme cardiaque, respiration. . .). D'autre part le système nerveux **parasympathique** qui s'activera lors de périodes de repos pour réduire l'activité et restaurer l'homéostasie viscérale ou la régulation de l'activité sexuelle (McCorry, 2007). L'essentiel de l'innervation de ce dernier système provient en réalité d'un seul nerf, le **nerf vague**, parfois aussi appelé "nerf pneumogastrique", qui est le dixième nerf crânien et aussi le plus étendu (du latin *vagus* : errant, vagant). Il est en particulier la principale innervation efférente du cœur. La stimulation directe de ce nerf cause une réduction

de la fréquence cardiaque, voire un arrêt complet du cœur. Cette influence du système parasympathique peut s'exercer extrêmement rapidement (en moins d'une seconde) mais celle-ci ne dure généralement que quelques battements. A l'inverse, le système sympathique mettra plus de temps avant d'accélérer complètement la fréquence cardiaque (au moins 5 secondes) mais cette influence pourra perdurer même plusieurs dizaines de secondes après la disparition du stimulus (Nunan et al., 2010).

L'importance de cette fonction inhibitrice efférente (provenant du cerveau), et son rôle dans la construction des états mentaux associés aux émotions a été soulignée au milieu du XIX^{ème} siècle par la publication du physiologiste français Claude Bernard d'une *Étude sur la physiologie du cœur* (Bernard, 1865), article portant sur les interactions entre la physiologie cardiaque et la construction des expériences émotionnelles. De façon intéressante, ce texte est intervenu dans un contexte scientifique où les connaissances de la physiologie naissante relativement à la fonction du muscle cardiaque cohabitaient avec un usage même du mot "cœur" fortement marqué par le sceau du romantisme et associé - sans doute encore plus qu'aujourd'hui - au centre de gravité de l'expression des passions de l'âme. La tendance était alors à la correction de cette vision aristotélicienne qui plaçait les passions dans le cœur au détriment de la fonction du cerveau renvoyé à un rôle de régulation thermique. Pourtant, le père de la médecine expérimentale voulut ici argumenter en la faveur d'une autre hypothèse : celle d'une influence mutuelle entre le cœur et le cerveau dans la construction des émotions, dont le muscle cardiaque ne pouvait pas être définitivement exclu. Il chercha alors à justifier la pertinence de cet usage pour "*expliquer par la physiologie ce qui n'a été jusqu'à présent qu'une simple intuition de l'esprit*". Il s'arrêta en particulier sur l'observation d'une inhibition temporaire du muscle cardiaque après la stimulation certains nerfs. S'il n'envisageait pas que la volonté seule puisse en faire cesser l'activité, il considérait en revanche que certaines sensations à l'origine des émotions puissent générer "*une solidarité d'actions réciproques des plus intimes*" où une accélération momentanée pouvait induire une diminution brusque de la fréquence par réflexe nerveux, réduisant ainsi le débit sanguin dans le cortex et amorçant, selon le physiologiste, les composantes subjectives de la négativité de l'émotion. On retrouve peu de temps après l'influence de cette proposition dans les textes de Charles Darwin qui, lorsqu'il traita de ces questions dans *l'Expression des émotions chez l'homme et les animaux* (Darwin, 1872), mentionnera : "*Un autre fait que je dois rappeler et, sur lequel Claude Bernard a insisté à plusieurs reprises, c'est que, lorsque le cœur est impressionné, il réagit sur le cerveau; l'état du cerveau réagit à son tour sur le cœur par l'intermédiaire du nerf pneumogastrique [i.e. nerf vague] ; en sorte que, sous l'influence d'une excitation quelconque, il se produit des actions et des réactions réciproques multipliées entre ces deux organes, les plus importants de l'économie*" (p.72 de la traduction française de 1874 par C. Reinwald). Ces premières observations ont été largement reprises par la suite et le développement de méthodes d'enregistrements précises de l'activité cardiaque (Billman, 2011) a alimenté l'idée selon laquelle les fluctuations

de la fréquence de battements du cœur pouvaient refléter plus que l'activité d'une pompe sanguine et décrire la nature des états mentaux éprouvés par l'individu, mais également les tentatives de régulation ou de contrôles qu'il peut tenter d'exercer inconsciemment sur eux.

La variabilité de la fréquence cardiaque

La fréquence cardiaque est en permanence adaptée aux besoins de l'environnement. La détection de stimuli induisant des réactions émotionnelles est un facteur important de cette influence. La variabilité de la fréquence cardiaque (*Heart Rate Variability*, HRV) correspond aux différences entre des intervalles successifs observés entre deux battements du cœur. Chez un individu en bonne santé, ces changements sont le reflet direct de l'influence conjointe des systèmes sympathique et parasympathique. Mais de façon déterminante, on observe également que la fréquence instantanée ne reste pas stable dans le temps. Le cœur n'est pas un métronome (Shaffer et al., 2014), mais présente au contraire des patterns qui en font varier le rythme de façon cyclique. Ces fluctuations peuvent être de très courte durée (d'un battement à l'autre) ou sur de très longues durées (suivant par exemple le rythme circadien). Des fluctuations particulièrement importantes s'observent sur des cycles d'une durée de 2 à 8 secondes (0.15-0.4 Hz, aussi qualifié de haute fréquence de la variabilité cardiaque, HF-HRV), ce qui correspond environ à un cycle de respiration. La fréquence cardiaque a tendance à légèrement augmenter lors de l'inspiration pour favoriser la distribution de l'oxygène et à décélérer lors de l'expiration, ce qui est également qualifié d'arythmie respiratoire sinusale (RSA). Nous avons évoqué l'influence inhibitrice que le nerf vagal peut avoir sur la fréquence cardiaque à très court terme. Comme, après une stimulation, le rythme retourne ensuite rapidement à la normale, une augmentation brusque de la fréquence cardiaque peut s'observer après une réduction de l'influence vagale sur le cœur, ce qui peut se produire lors de phénomènes de surprise où une augmentation très rapide est nécessaire. Mais dans les deux cas, qu'il s'agisse d'une brusque accélération ou d'une brusque décélération, l'influence vagale est à l'origine de ces fluctuations. Par ailleurs, comme l'activité du système nerveux parasympathique est dominante dans les périodes de repos (e.g. position assise), on considère que l'arythmie respiratoire sinusale, correspondant aux "hautes fréquences" de la variabilité cardiaque, est un reflet fiable de l'influence vagale sur le cœur (Akselrod et al., 1981) et donc une des principales influences du cerveau sur celui-ci.

Cette variabilité de la fréquence cardiaque - les changements d'intervalles d'un battement à l'autre - représente quelque chose de très différent de la fréquence cardiaque - instantanée ou plus généralement moyennée sur de plus grandes périodes de temps. L'importance clinique de cette mesure a été soulignée au milieu des années 1960 lorsque (Hon and Lee, 1963) ont montré que l'hypoxie fœtale (la diminution de l'oxygénation du fœtus) est précédée par des changements dans la variabilité de la fréquence cardiaque avant même qu'il y ait des changements observables dans la fréquence cardiaque générale. Ces observations préliminaires, qui concernaient alors

des aspects essentiellement médicaux, ont été généralisées et il a été établi qu'une réduction de cette variabilité cardiaque apparaissait en lien avec le diabète (Vinik et al., 2003), le tabagisme (Karakaya et al., 2007), l'obésité (Skrapari et al., 2007), l'hypertension (Pagani et al., 1984) les problèmes cardio-vasculaires (De and Randall, 2005), et était un prédicteur robuste de la mortalité - toute causes confondues - parmi des populations âgées (Tsuji et al., 1994).

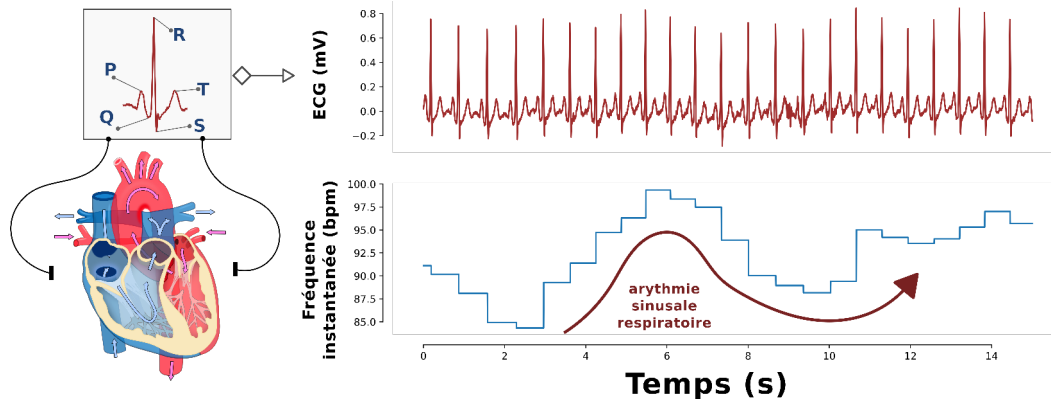


Fig. 2.1: Variabilité de la fréquence cardiaque. La mesure de la fréquence cardiaque est réalisée lors d'une électrocardiographie (ECG). La représentation graphique qui en résulte (l'électrocardiogramme) reflète l'activité électrique du cœur, ici représentée lors d'un seul battement. Cette représentation varie en fonction du nombre et du placement des dérivation de l'ECG. On observe un ensemble d'inflexions positives et négatives : P-Q-R-S-T. Le complexe QRS correspond à la dépolarisation et à la contraction des ventricules droit et gauche. La fréquence cardiaque se mesure via les intervalles séparant deux ondes R. Deux battements suffisent pour extraire une fréquence instantanée, représentée en bleu et qui s'exprime en battements par minutes (bpm). La variabilité de la fréquence cardiaque, et notamment l'arythmie sinusale respiratoire, s'observera à travers les variation de cette fréquence instantanée dans le temps.

Ces résultats ont permis d'en faire un indicateur robuste de l'état de santé physique. Mais, il est également apparu que cet indice était également affecté par des états mentaux ou des profils psychopathologiques particuliers et se trouvait influencé par des processus ayant une origine bien plus élevée dans la hiérarchie corticale et représenterait les capacités d'adaptation de l'organisme à la variabilité de son environnement. Cette influence a participé à redéfinir les interactions sous-estimées entre le système nerveux central - et plus particulièrement le cortex - et différentes composantes du système nerveux autonome. Ces révisions ont été en particulier synthétisées par deux théories : la **théorie polyvagale** développée par le physiologiste Stephen Porges (Porges, 1995), et le **modèle d'intégration neuroviscéral** proposé par le neuroscientifique Julian Thayer (Thayer and Lane, 2000). Comme nous allons le voir, ces modèles ont l'avantage d'intégrer le développement des connaissances anatomiques des différents systèmes nerveux avec leur influence sur les composantes comportementales et cognitives des émotions à travers une perspective évolutive concrète.

La théorie polyvagale

Les remarques de Claude Bernard, et leur commentaire plus tard par Darwin, avaient insisté sur l'interaction et la dynamique qu'entretenaient certainement le cœur et le cerveau dans la production des états émotionnels. Pourtant, la description plus systématique du système nerveux autonome qui sera proposée plus tard par (Langley, 1921) a participé à minimiser le rôle des structures de régulation et des signaux afférents dans l'économie de ce système, au profit d'une distinction plus marquée entre les subdivisions sympathiques et parasympathiques. Cette vision a participé au raffinement des connaissances sur ces deux systèmes, mais a eu pour conséquence de considérablement éclipser les interactions possibles entre le système autonome et le système nerveux central et à considérer que les manifestations physiologiques pouvaient se comprendre essentiellement par cette double influence.

La théorie polyvagale (Porges, 2009; Porges, 2011; Porges, 2007) proposée au milieu des années 1990 par le physiologiste Stephen Porges est venue intégrer un ensemble de résultats suggérant que cette dichotomie était insuffisante pour expliquer l'ensemble des comportements des mammifères, et notamment les comportements sociaux. Porges s'est notamment intéressé à un ensemble de résultats montrant que le nerf vague n'est pas un système unitaire mais est constitué de sous-composantes ("poly"-“vagale”) apparues lors de différentes périodes de l'évolution et sous-tendant des comportements différents. En plus du système autonome sympathique initiant les comportements de fuite ou de lutte, la composante parasympathique peut à nouveau être divisée en deux. Une voie non myélinisée partant du noyau moteur dorsal du nerf vague et innervant les organes situés au-dessous du diaphragme est la composante la plus ancienne. Elle est commune à l'ensemble des vertébrés et se trouve impliquée dans la digestion mais aussi dans la réponse à la menace par une réduction de la fréquence cardiaque et permet les comportements d'immobilisation (feindre la mort, se terrer pour échapper à un prédateur) qui sont des réflexes défensifs anciens dans l'évolution. Une seconde voie, celle-ci myélinisée et permettant donc une communication rapide de ses influx nerveux, part du noyau ambigu et innerve les organes situés au-dessus du diaphragme, dont le cœur. Cette dernière est propre aux mammifères (Beauchaine et al., 2007) et permettrait une régulation rapide de l'activité cardiaque en fonction du niveau d'engagement et d'attention envers l'environnement et intervient dans les comportements sociaux. De plus, les noyaux du tronc cérébral impliqués dans la régulation de cette composante myélinisée du nerf vague se sont intégrés au cours de l'évolution aux noyaux impliqués dans le contrôle des muscles faciaux (orientation des yeux, expressions faciales et capacités de communication), suggérant une association forte entre les comportements sociaux et le contrôle des états physiologiques du corps (Beauchaine et al., 2007; Porges, 2001; Porges, 2003).

L'une des particularités de cette théorie est qu'elle place l'engagement avec autrui au centre des processus de santé physique et mentale. En particulier, la variabilité de la fréquence cardiaque telle qu'elle s'observe à travers l'arythmie sinusale respiratoire serait un reflet de la composante la plus récente du nerf vague, ce qui fait de cette

mesure un indice de la régulation du système nerveux autonome parasympathique par des structures corticales plus récentes dans l'évolution.

Le modèle d'intégration neuroviscéral

Au début des années 2000, un second modèle est venu compléter ce que la théorie polyvagale avait participé à mettre en lumière. Le modèle d'intégration neuroviscéral (Thayer and Lane, 2000; Thayer et al., 2009; Smith et al., 2017) a poursuivi l'étude des facteurs influençant la variabilité cardiaque en se concentrant particulièrement sur la hiérarchie corticale impliquée chez l'homme. Ce modèle suggère en particulier que les réseaux neuronaux engagés dans le contrôle inhibiteur et la régulation émotionnelle, notamment la partie droite du cortex préfrontal controlatéral, sont également impliqués dans le contrôle de l'activité cardiaque via l'influence inhibitrice du nerf vague. De nombreuses observations sont venues alimenter cette thèse, en suggérant notamment que les indices de variabilité de la fréquence cardiaque, en particulier lorsqu'ils sont mesurés lors de périodes de repos de plusieurs minutes, sont corrélés avec des mesures comportementales, cognitives ou psychopathologiques classiquement utilisées pour quantifier les capacités de régulation émotionnelle, d'attention ou encore de contrôle inhibiteur.

La variabilité au repos de la fréquence cardiaque a ainsi été associée avec la connectivité structurelle (Wei et al., 2018) et fonctionnelle (Sakaki et al., 2016) entre l'amygdale et le cortex préfrontal, les capacités attentionnelles (Gazzellini et al., 2016), le contrôle inhibiteur dans le cadre de la tâche de "*Think/No-Think*" (Gillie et al., 2013; Gillie et al., 2015) ou encore les capacités de théorie de l'esprit (Lischke et al., 2017). Sur le plan clinique, une réduction de la variabilité de la fréquence cardiaque a été observée dans le cadre de troubles anxieux (Chalmers et al., 2014) ainsi que dans le syndrome de stress post-traumatique (Gillie and Thayer, 2014) où la variabilité cardiaque participe directement à prédire la prédisposition du trouble (Minassian et al., 2015), son apparition (Park et al., 2017) mais aussi la résilience des symptômes (Shah et al., 2013). Les parties latérales du cortex préfrontal sont considérées comme les structures apicales de la hiérarchie neurale interagissant avec la composante parasympathique du système nerveux, hiérarchie comprenant 8 niveaux dans la version la plus récente de ce modèle (Smith et al., 2017).

2.2 Étude n°1 - Modulation de l'activité cardiaque induite par la suppression volontaire de souvenirs à valence négative

Dans leur ensemble, la théorie polyvagale et le modèle d'intégration neuroviscéral suggèrent l'existence de connexions directes entre les parties les plus récentes du cortex sous-tendant les comportements sociaux, les capacités attentionnelles et inhibitrices, et les circuits vagues du système parasympathique engagés dans la régulation du rythme cardiaque. Alors même que des études ont montré des liens entre les profils de variabilité cardiaque et la capacité de suppression mnésique lors du paradigme "Think/No-Think" (Gillie et al., 2013; Gillie et al., 2015), l'engagement direct de tels circuits dans la régulation émotionnelle que pourrait permettre le contrôle inhibiteur des souvenirs reste encore très mal défini. La démonstration d'un lien de cette nature serait pourtant une étape importante dans la compréhension des mécanismes de régulation émotionnelle.

Il est souvent objecté à la stratégie de suppression directe des souvenirs qu'un tel procédé pourrait être associé sur le plan clinique à un effet de retour, parfois théorisé dans un cadre psychanalytique, en postulant notamment que ces représentations ne disparaissent pas réellement de la mémoire mais peuvent ressurgir par la suite avec un contenu émotionnel renforcé. La seconde objection qui peut être faite se situe davantage sur le plan théorique et souligne l'ancrage profondément physiologique des émotions qui devraient être supprimées, et la grande difficulté, voire l'impossibilité, pour le contrôle inhibiteur mis en place par le cortex préfrontal, d'atteindre les réseaux du système nerveux autonome, laissant les empreintes corporelle des émotions inchangées.

La première étude que nous présentons ici apporte des arguments contre ces deux lignes de critique et suggère que les capacités cognitives de contrôle sur les souvenirs épisodiques doivent être envisagées dans un cadre bien plus incarné qu'auparavant tenant compte de la dimension polyvagale et neuroviscérale des réseaux de contrôle. Ce travail a été conduit sur deux groupes de participants volontaires, jeunes (18-35 ans) et ne présentant pas de troubles psychiatriques identifiés. Nous avons exploré les conséquences physiologiques de la suppression volontaire de souvenirs via le paradigme "Think/No-Think" à l'aide d'enregistrements d'électrocardiographie. Nous rapportons ici les résultats tels qu'ils ont été présentés, à l'exception de quelques modifications syntaxiques mineures, dans la version en préimpression n°1 de l'article intitulé "Does the heart forget? Modulation of cardiac activity induced by inhibitory control over emotional memories" qui est disponible à l'adresse suivante : <https://www.biorxiv.org/content/early/2018/07/25/376954>.

Does the heart forget? Modulation of cardiac activity induced by inhibitory control over emotional memories

Nicolas Legrand¹, Olivier Etard^{2,3}, Anaïs Vandeveld³, Mélissa Pierre¹, Fausto Viader¹, Patrice Clochon¹, Franck Doidy¹, Denis Peschanski⁴, Francis Eustache¹, and Pierre Gagnepain¹

¹Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France

²CHU de Caen, Service des explorations fonctionnelles du système nerveux, Caen, F-14000, France

³Imagerie et Stratégies Thérapeutiques de la Schizophrénie (ISTS), Normandie Univ, UNICAEN, Faculté de médecine, CYCERON, Caen F-14000, France

⁴European Centre for Sociology and Political Science (Université Paris 1, EHESS and CNRS), Paris, France

Abstract

The subjective construction surrounding the perception of negative experiences is partly built upon bodily afferent information, comprising heart, gut or respiratory signals. While this bottom-up influence has been extensively described, the opposite pathway, the putative influence of cognitive processes over autonomic response, is still debatable. However, emotion regulation and the ability to control maladjusted physiological response associated with thoughts and memories is a recurrent concern for most psychiatric disorders and mental health as a whole. Memory suppression (i.e. exerting inhibitory control toward unwanted memories) has been proposed as a possible solution to perform such regulation. However, this method also holds debates as it could putatively worsen distressing reactions when unsuccessful, or simply be ineffective on the physiological roots of emotions. Here, we tested the hypothesis that memory suppression can influence the physiological marker of emotions in two studies using the “Think/No-Think” paradigm. We measured in healthy participants the aftereffect of suppressing negative memories on their cardiac reaction toward the excluded memories. Results of Study 1 revealed that an efficient control of memories was associated with a long-term inhibition of the cardiac deceleration normally induced by disgusting stimuli. Study 2 replicated this finding and critically showed that it was linked to the decrease of the 5-9 Hz frequency band, a prominent electrophysiological marker of memory suppression. These results support the notion that cognitive control over unwanted emotional memories can influence in parallel autonomic processes to achieve emotional regulation, and open new avenues for possible interventions focused on training this suppression mechanism to reduce the impact of distressing intrusive memories on mental health.

Keywords: *forgetting, emotion regulation, memory suppression, heart rate variability, theta*

Introduction

Intrusive memories often take the form of distressing images and physical reactions that interrupt ongoing mentation (Brewin, Gregory, Lipton, & Burgess, 2010). Difficulties to control those inappropriate distressing emotional responses, which may follow or initiate intrusive mental images, can lead to dangerous or harmful behaviors (Gross, 2015; Etkin, Büchel, & Gross, 2015), as is seen in numerous psychiatric conditions (Aldao, Nolen-Hoeksema, & Schweizer, 2010; Fernandez, Jazaieri, & Gross, 2016; Berking & Wupperman, 2012). Various symptoms of anxiety-related disorders comprising worries, the recall of traumatic experiences in PTSD (Shepherd & Wild, 2014) or unwanted thoughts in obsessive-compulsive disorders (Fergus & Bardeen, 2014), as well as ruminations observed in depression (Jormann & Gotlib, 2010; Mihailova & Jobson, 2018) are representative of such difficulties. Due to historical presumptions about the independent and pernicious persistence of repressed memories, it is often assumed that attempt to exclude from awareness those distressing im-

ages is harmful (Wegner & Zanakos, 1994; Dunn, Billotti, Murphy, & Dalgleish, 2009). By this view, repressed memories may backfire, causing psychological symptoms (Moritz et al., 2014; Rachman, 2007; Ólafsson et al., 2014), or possibly aggravating autonomic responses that accentuate perceived distress (Campbell-Sills, Barlow, Brown, & Hofmann, 2006; Gross, 1998). Alternatively, the origins of excessive autonomic and distressing reaction linked with intrusive memories may arise not because suppression is intrinsically bad, but because the cognitive computations achieved by the memory control network to deploy the necessary functional coping skills are disrupted or inappropriate. Here, we examine whether stopping episodic retrieval to suppress an unpleasant image might also induce additional modulation of heart rate activity.

Although targeting different components, the control of intrusions and emotions is believed to be underpinned by common neurocognitive processes orchestrated by the prefrontal cortex (Depue, Orr, Smolker, Naaz, & Banich, 2016), acting protectively like an “*immune system of the mind*” (Cole, Repovš, & Anticevic, 2014). Suppressing memory retrieval has been consistently reported to engage an inhibitory modulation of areas supporting memory retrieval such as the hippocampus (Anderson, 2004; Depue, Curran, & Banich, 2007) or the visual cortex (Gagnepain, Henson, & Anderson, 2014). This top-down suppression is believed to originate primarily from the right middle frontal gyrus (MFG) (Benoit & Anderson, 2012; Benoit, Hulbert, Huddleston, & Anderson, 2015). Critically, a similar top-down modulation is also observed for the regulation of emotional response to negative stimuli (Eippert et al., 2007; Hayes et al., 2010; Kim & Hamann, 2007; Ochsner et al., 2004), with a specific target on the amygdala (Kohn et al., 2014; Comte et al., 2016). Comparisons of retrieval suppression and affect regulation show an overlapping localization within the right MFG (Depue, Orr, Smolker, Naaz, & Banich, 2015). Critically, suppressing retrieval of aversive images without any attempt to regulate emotion still reduces hippocampal and amygdala activity (Depue, Burgess, Willcutt, Ruzic, & Banich, 2010; Gagnepain, Hulbert, & Anderson, 2017; Depue, Curran, & Banich, 2007), suggesting that suppression may regulate both memory and affect during unpleasant intrusions. Confirming this idea, (Gagnepain, Hulbert, & Anderson, 2017) showed that people’s memory suppression skills in engaging the right frontoparietal network to downregulate both hippocampal and amygdala responses during intrusive memory was related to a later reduction of their perceived emotional state.

Together, these results suggest that inhibitory control, the ability to stop or override a

specific response comprising distressing thoughts and emotional experiences, may have strong clinical relevance in the context of various psychiatric disorders. Notably, if previous works suggest a possible beneficial effect on the cognitive and experiential level of emotions, this hypothesis questions the influence of cognitive control over other central characteristics of emotion and mental distress: being a set of automated physiological reactions. In the late 19th century, the psychologist and philosopher William James referred to emotions as feelings accompanying bodily changes (James, 1884). Several researches have since supported and completed this broad framework of embodiment, positing bodily afferent signals as the roots of emotional construction (Damasio & Carvalho, 2013; Craig, 2002; Garfinkel & Critchley, 2016; Babo-Rebello, Richter, & Tallon-Baudry, 2016; Pezzulo et al., 2018).

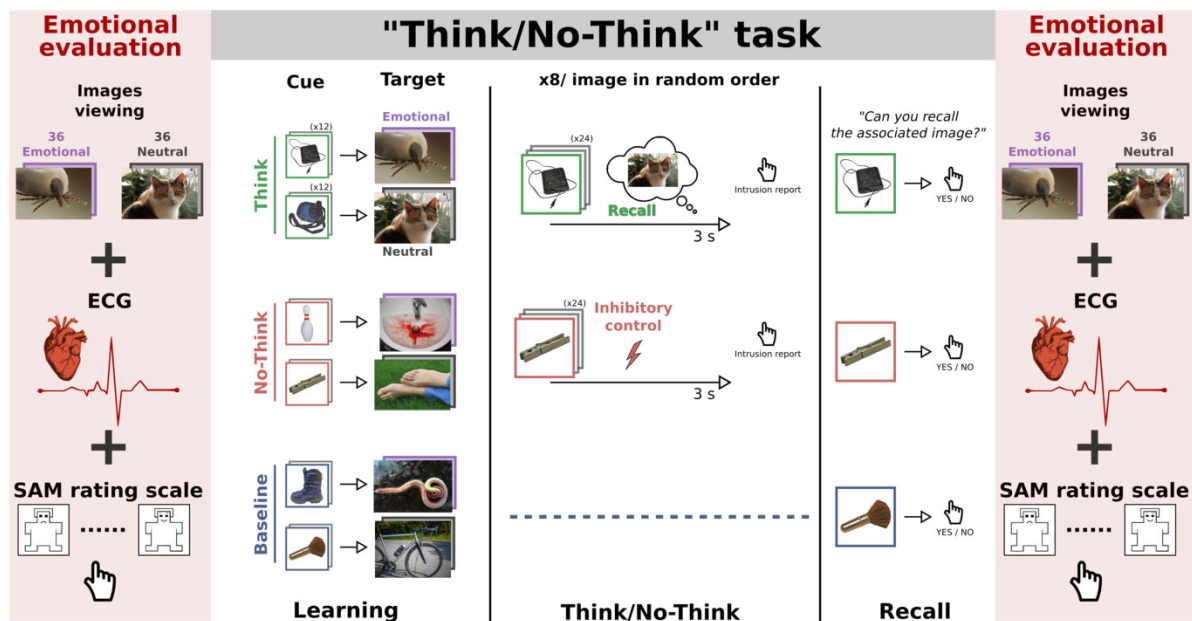
From a physiological perspective, the activation of the autonomic nervous system (ANS) and its influence on cardiac activity is one of the most prominent manifestations of emotional reactions (Kreibig, 2010; Ekman, Levenson, & Friesen, 1983). Indeed, changes in the duration between successive heartbeats, referred as the *heart rate variability* (HRV), is representative of both sympathetic and parasympathetic influences reflecting affective responses (Valenza, Citi, Lanatá, Scilingo, & Barbieri, 2014; Ekman, Levenson, & Friesen, 1983; Critchley et al., 2005). However, the role of the prefrontal control network during cardiac modulation and its relation to the neurocognitive processes supporting the regulation of memory and emotions remains unclear. Here, we tested the hypothesis that the direct suppression of emotional memories not only affect their cognitive and phenomenological dimensions but also inhibit later cardiac reactions.

This hypothesis is motivated by two distinct lines of findings. The first line of findings is based on the Think/No-Think (TNT) paradigm (Anderson & Green, 2001) which measures the ability of the participants to control and suppress awareness of previously encoded memories. Using this procedure, (Gagnepain, Hulbert, & Anderson, 2017) recently demonstrated that the voluntary suppression of aversive and unwanted memories was associated with a greater down-regulation of activity in both the amygdala and the hippocampus. Critically, however, reduced amygdala activity was linked to a reduction of the perceived valence associated with the excluded content, and was not merely an indirect consequence of hippocampal suppression; rather, both the hippocampus and the amygdala were targeted by a top-down inhibitory control signal originating from the right MFG. The second line of evidence is based on the possible influence of prefrontal areas on autonomic processes. This notion has been popularized in the *neurovisceral integration model* (Smith, Thayer, Khalsa, & Lane, 2017; Thayer &

Lane, 2000; Thayer & Lane, 2009) which proposes that cardiac vagal control is orchestrated by higher-order brain areas in the cortical hierarchy, such as the dorsolateral (Åhs, Sollers, Furmark, Fredrikson, & Thayer, 2009; Napadow et al., 2008; Smith, Allen, Thayer, Fort, & Lane, 2014; Winkelmann et al., 2016; Critchley et al., 2005) and medial (Adhikari et al., 2015; Sakaki et al., 2016; Lane et al., 2009) prefrontal areas, whose activity might regulate the central nucleus of the amygdala. In this model, the amygdala is a major efferent source of modulation of cardiovascular response via the (parasympathetic) vagal nerve. The poly-vagal theory (Porges, 1995; Porges, 2001; Porges, 2011) further proposes that efferent vagal connections supporting parasympathetic deceleratory signals to heart activity may be exerted by two different pathways following a phylogenetic hierarchy. One older dorsal branch is associated with primal strategies of freezing and metabolic resources preservation that both require heart rate deceleration. Another more recent ventral branch, which may be distinctly mammalian, is myelinated and provide both inhibitory and dis-inhibitory control over limbic circuits in the service of social behaviors. This would suggest that this more recent “smart vagus” branch could selectively modulate heart activity depending on the context, thus integrating attention, emotion or memory inputs in support of autonomic regulations. Giving these two lines of evidence, we propose that the cognitive control network engaged to regulate the mnemonic and emotional content of unwanted memory traces through the suppression of hippocampus and amygdala activity, respectively, may also inhibit later cardiac responses to suppressed stimuli.

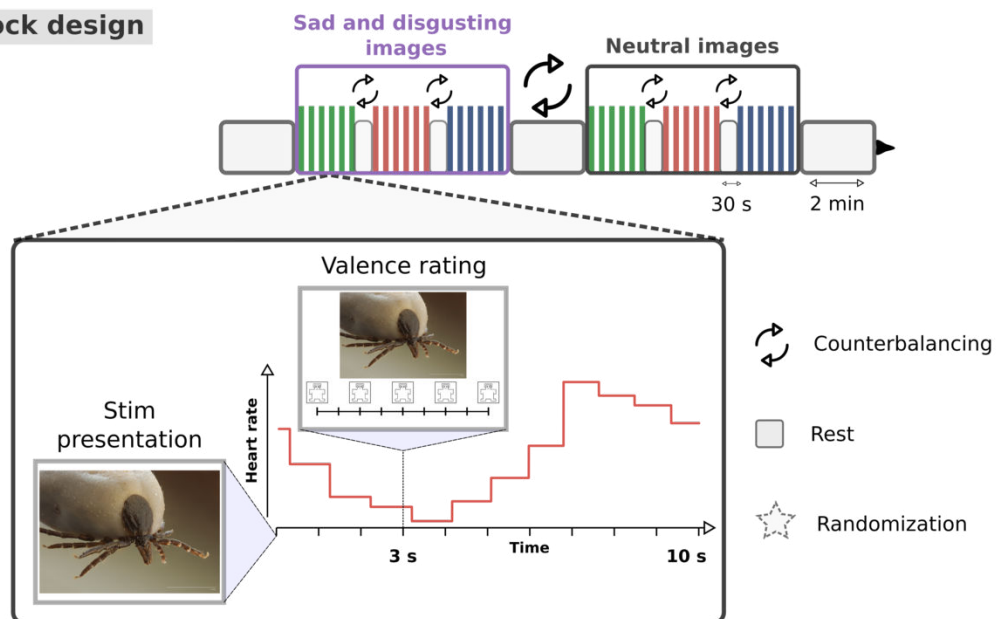
To test this hypothesis, we measured in two consecutive studies the electrocardiographic (ECG) response to negative and neutral scenes before and after the TNT paradigm (Fig. 1). During the TNT task, participants attended to a reminder (an object) of these emotional and neutral scenes which cued them to retrieve the scene (Think items), or to suppress its retrieval (No-Think items). During pre- and post-TNT emotional evaluation, participants rated the Think and No-Think scenes’ valence, along with previously studied Baseline scenes not presented during this TNT phase. A cued-recall task was also performed immediately after the TNT task to estimate the aftereffects of memory suppression on recall performances. This procedure was applied in Study 1 using a blocked presentation of Think, No-Think, and Baseline pictures (Fig. 1) during the emotional evaluation task, and two emotions: disgust and sadness. Study 2 focused on emotional disgust, employed a random design instead of a block one, and additionally recorded concurrent electroencephalographic (EEG) response during the TNT task to track the oscillatory dynamics supporting the inhibitory control of memories and estimate their role in the inhibition of the later cardiac response.

A. Experimental phases



B. Pre and post-TNT emotional evaluation

Study 1: block design



Study 2: randomized design



Figure 1: Experimental phases and emotional evaluations.

Legend Fig. 1: A. Participants learned object-scenes associations. The associated scene had either neutral or negative valence (Study 1: 36 neutral, 18 sad and 18 disgusting; Study 2: 36 neutral and 36 disgusting). After this procedure, they underwent the TNT task. The cue from Think and No-Think condition was presented on the screen bounded with either green or red frame respectively. Participants had to recall the associated scene if the frame was green and to prevent the memory from entering awareness if the frame was red. The 48 cues were presented in random order and this procedure was repeated 8 times. The baseline items were not presented during the TNT task. During the final recall, all the cues were presented on the screen and participant had to report if they can or not recall the associated scenes, and then a detailed description of the scene if they could so. The difference of correct recall between No-Think and Baseline provided an estimate of memory suppression performances. **B.** We measured subjective and physiological reaction toward the 72 scenes before and after the TNT procedure (pre-TNT and post-TNT evaluations). For both studies, neutral and negative images were presented using a block design interleaved with resting periods to control for the long-lasting effect of emotion on the autonomic response. Inside each block, we presented the Think, No-Think and Baseline items either with a block design (Study 1) or using a randomized order (Study 2). Each scene was presented on full-screen for 3 seconds, after that a SAM valence rating scale appeared under a reduced version of the picture and participants had to report their affective rating of the scene. Heart rate was recorded during the whole procedure. We illustrate the valence rating task for a single trial presentation with the corresponding heart rate activity. The object illustrating the memory cue are from the BOSS database and the pictures illustrating the paired scenes are free-use pictures taken from the internet for illustrative purpose.

Results Study 1

Behavioral results

Intrusions When successful, the inhibitory control of memory recall is characterized by a progressive decrease in intrusion frequency (Levy & Anderson, 2012; van Schie & Anderson, 2017; Hellerstedt, Johansson, & Anderson, 2016; Hu, Bergström, Gagnepain, & Anderson, 2017; Benoit, Hulbert, Huddleston, & Anderson, 2015). Here, we compared the frequency of intrusion of Disgust, Sadness and Neutral items over the 8 TNT blocks for the No-Think

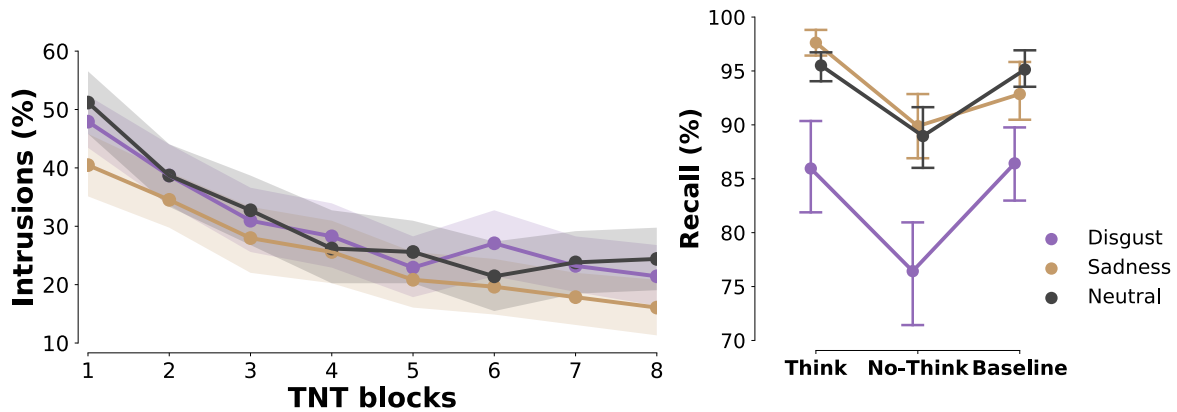
condition (Fig. 2 A). An Emotion \times TNT blocks ANOVA showed both an effect of Emotion ($F_{(1.60, 43.18)} = 3.82, \eta_g^2 = 0.007, p = 0.04$) and TNT blocks ($F_{(3.21, 86.62)} = 21.00, \eta_g^2 = 0.08, p < 0.001$) but no interaction between these two factors ($F_{(7.41, 200.10)} = 0.78, \eta_g^2 = 0.003, p = 0.62$). Planned comparisons revealed that participants had on average less intrusions associated with disgusting images as compared to both neutral ($t_{(27)} = 2.45, p = 0.02, d = 0.46$, two-tailed) and sad stimuli ($t_{(27)} = 2.04, p = 0.05, d = 0.38$, d, two-tailed).

Inhibition of recall induces forgetting. Preventing memory from entering awareness during the TNT task is associated with a higher forgetting of the suppressed scenes during the post-TNT recall task, compared to baseline items (Anderson & Green, 2001). We first controlled for the successful suppression of No-Think items by examining the percentage of correctly recalled items following the Think/No-Think task. We sought to replicate the standard finding associated with the inhibition of involuntarily recall, and compared the percentage of recall of No-Think items to Baseline items. Baseline items had been trained in the same initial learning session with the Think and No-Think items but were omitted from the TNT phase. They provide a baseline estimate of memorability (or emotionality) of the scenes, given that neither suppression nor retrieval had been performed on them. These pairs enabled us to assess the effects of retrieval and suppression on the retention or emotional judgment of Think and No-Think items respectively, against pairs that were similarly old and trained (Fig. 2 B). An Emotion \times Condition [Think, No-Think, Baseline] ANOVA on recall performances (percentage of correctly recalled scenes – see Method) revealed a significant effect of both Emotion ($F_{(1.32, 35.51)} = 16.01, \eta_g^2 = 0.09, p < 0.001$) and Condition ($F_{(1.74, 47.04)} = 8.17, \eta_g^2 = 0.05, p = 0.001$), but no interaction between those two factors ($F_{(2, 61, 70.44)} = 0.86, \eta_g^2 = 0.007, p = 0.45$). Planned comparisons showed that participants recalled significantly less No-Think items than Baseline for Disgust ($t_{(27)} = -2.27, p = 0.03, d = -0.43$) and Neutral condition ($t_{(27)} = -3.12, p = 0.004, d = -0.58$) while no difference was found for Sadness ($t_{(27)} = -0.96, p = 0.34, d = -0.18$).

Taken together, these findings are inconsistent with the notion that negative stimuli may be more difficult to control and are in line with previous works reporting numerically less frequent intrusions (Gagnepain, Hulbert, & Anderson, 2017) and greater memory suppression (Lambert, Good, & Kirk, 2010; Depue, Curran, & Banich, 2007) for negative memories. On one hand, this pattern of behavioral findings can reflect the existence of a greater motivational force driving the need to control and suppress memory for disgusting pictures. On the other hand, it may also reflect the greater natural need to avoid disgusting pictures and

reduce their storage (Medina, Clark, & Thorne, 2016; Newhagen, 1998), which is compatible with the fact that we observed an overall poorer recall of disgusting stimuli compared to sad or neutral scenes. However, although we cannot tease apart those two hypotheses in the current study, it is worth mentioning that they are not mutually exclusives and both rooted from the natural desire to avoid and reject emotional disgust.

A. Intrusion proportion across TNT blocks **B. Total recall after the TNT task**



C. Difference in valence rating between pre and post-TNT evaluation

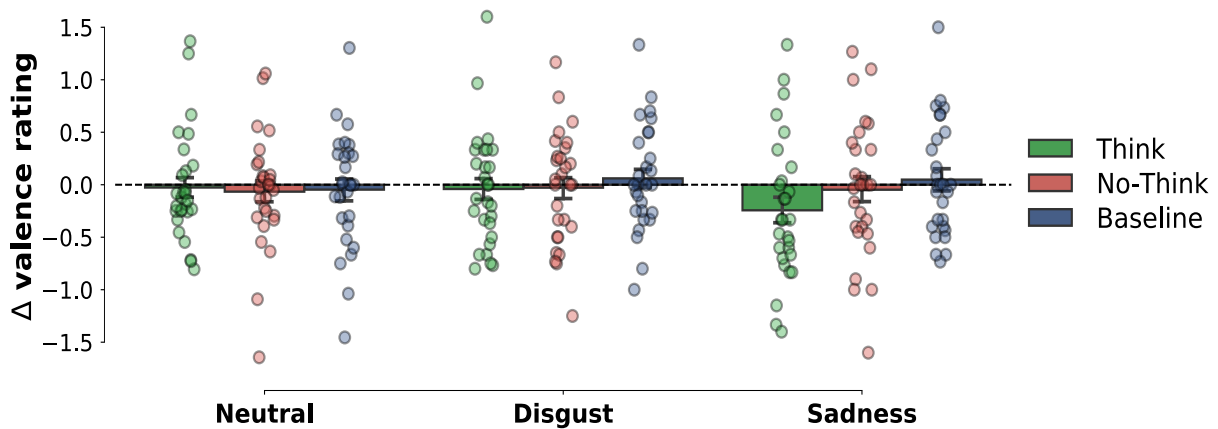


Figure 2: Behavioral indices of inhibitory control over memory and emotion in Study 1 (n=28).

Legend Fig. 2: A. Intrusions proportions for No-Think trials (i.e. proportion of trials where the associated memory entered into awareness while participants were instructed to inhibit recall) over the eight TNT blocks. Participants increased their ability to control

intrusion over the eight TNT blocks. On average, disgusting pictures were reported less intrusive than sad or neutral ones. **B.** Total recall after the TNT procedure. Images in the suppression condition (No-Think) were more forgotten than baseline for both disgusting and neutral emotion. However this effect was not significant for sad stimuli. **C.** Difference in valence rating between pre- and post- TNT evaluation. We found no significant differences in term of subjective rating for the No-Think pictures as compared to baseline in any of the three emotional conditions. Shaded error bands and error bars represent the bootstrapped standard-error of the mean.

Effect of suppression in the subjective evaluation. Beyond this well-documented suppression of memory representation, we also questioned whether the ability to control intrusion could relate to changes in the subjective emotional judgment. We then examined the SAM valence rating after the Think/No-Think procedure adjusted by pre-TNT evaluations (Fig. 2 C). An Emotion \times Condition ANOVA revealed no significant effect of Emotion ($F_{(1.70, 45.87)} = 0.28, \eta_g^2 = 0.003, p = 0.72$), Condition ($F_{(1.95, 52.57)} = 1.17, \eta_g^2 = 0.008, p = 0.32$), or the interaction between those two factors ($F_{(3.63, 98.10)} = 0.93, \eta_g^2 = 0.01, p = 0.44$). Planned comparisons between No-Think and Baseline conditions revealed no significant difference for Disgust ($t_{(27)} = 0.62, p = 0.53, d = 0.18$), Sad ($t_{(27)} = 0.75, p = 0.45, d = 0.14$), or Neutral pictures ($t_{(27)} = 0.14, p = 0.88, d = 0.02$). Thus, suppression did not consistently affect the perceived valence of the scenes. An image's perceived valence would be expected to depend, however, on the context and the valence of surrounding pictures which may modulate decision processes and affective rating. The block procedure used for Study 1, bring together items from Think, No-Think, or Baseline conditions, and decision process may therefore be confounded by the context of their occurrence. Moreover, the binary yes/no intrusion scale used in study 1 may also have reduced the sensitivity and the awareness towards intrusions, reducing the engagement of adaptive control processes to purge momentary awareness and inhibition of the perceived valence. These potentials issues are addressed in Study 2.

Physiological results

Sadness and disgust emotions decrease heart rate. We then asked if repeatedly suppressing unwanted memories across blocks of the TNT task was accompanied by persisting modulation of the cardiac response that is normally elicited by the presentation of negative scenes. Before addressing this question, we first used the pre-TNT evaluation of emotional response to characterize and isolate the specific cardiac autonomic response elicited by ne-

gative affect and unpleasant scenes. The autonomic nervous system is a central marker of emotional elicitation (Kreibig, 2010). The influence of an emotion like disgust can be observed on the heart rate over long (e.g. > 1 min.) periods of exposure as during film presentation (Rohrmann & Hopp, 2008). However, given that our material only contains fixed images, the emotional effect is thought to be present on a shorter duration (between 2 and 8 seconds, corresponding to the necessary delay for the central nervous system to influence cardiovascular response), and to significantly decay after that. Moreover, the repeated presentation of long stimulation with a fixed time interval is thought to induce some synchronization observable on the respiratory sinus arrhythmia (RSA), which could thus add noise especially when averaging over long periods. Here, we analyzed the heart rate evoked variation for each scene up to 10 seconds after the presentation. This design and the fixed regular presentation of image every 10 seconds allowed to account for this synchronization and restrict the analysis to differences between emotions. Such short periods of time have been successfully used in previous works to characterize and quantify the heart autonomous response to emotion (Valenza, Citi, Lanatá, Scilingo, & Barbieri, 2014; Critchley et al., 2005). This method also offers the advantages to focus on the period preceding the emotional valence rating which reflects more precisely the item specific emotional content. In addition, this method also offers a greater flexibility to deal with spontaneous artifacts which are problematic in the context of a block analysis.

To assess the influence of an emotional elicitation on the following heart rate activity over the 10 seconds of presentation, we compared the averaged BPM curve across the three emotional conditions in the first pre-TNT evaluation (see Fig. 3 A). Results show that presenting scenes categorized as Disgusting or Sad induced a significantly greater deceleration of the heart rate than Neutral pictures. Neutral scenes induced a maximal deflection peak (MDP) of -0.4 bpm at 1.84 seconds after the stimuli presentation, while the deceleration was stronger both for Disgusting (MDP = -2.34 bpm after 3.44 seconds) and Sad stimuli (MDP = -2.24 bpm after 3.36 seconds). Altogether, emotional stimuli showed a significantly larger decrease of bpm as compared to Neutral ones with an emotional time of interest (TOI) ranging from 1.88 to 5.03 seconds. The peak of maximum difference observed at 3.43 seconds ($t_{(27)} = 4.59$, $P\text{-FDR} < .001$, $d = 0.81$, two-tailed paired t-test). This deceleratory pattern is similar to previous reports made using either facial expressions (Critchley et al., 2005) or films (Rohrmann & Hopp, 2008).

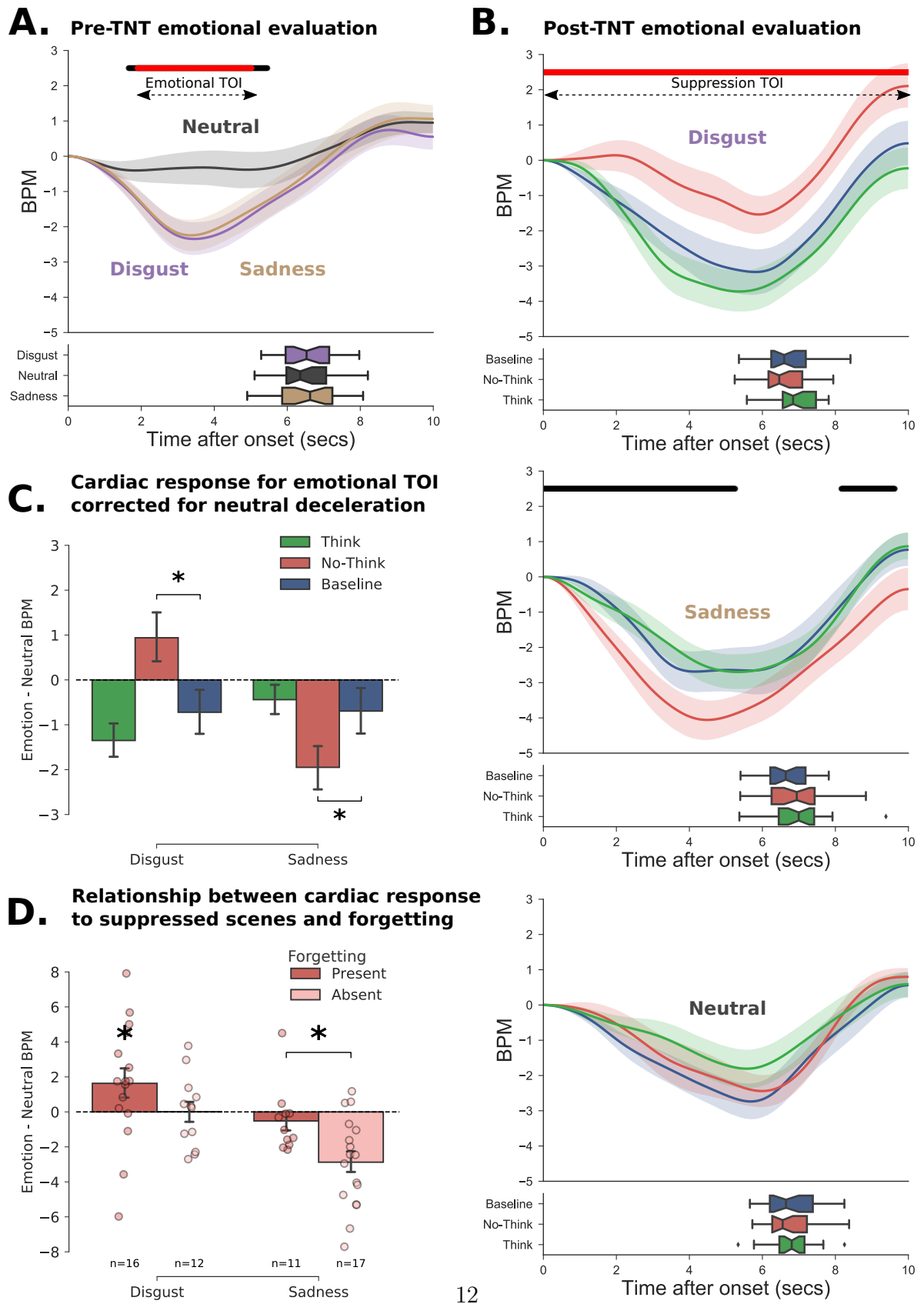


Figure 3: Cardiac response toward stimuli before and after the TNT procedure.

Legend Fig. 3: **A.** Cardiac response following the initial presentation of the images before the TNT task. The curves represent the changes in the number of heartbeat per minutes (BPM) using the onset of image presentation as a zero starting-point. In the lower panel, the boxplots represent the distribution of response times for the valence rating. **B.** Cardiac response after the TNT task for Disgust (top panel), Sadness (middle panel), and Neutral (bottom panel) images. For each emotion, we represent Think, No-Think and Baseline items separately (green, red and blue lines respectively). We observed a significantly reduced deceleration of the heart rate for disgusting No-Think items compared to baseline, while a greater deceleration was reported for sad picture in the suppression condition. For both panel A and B, the black line shows the significant difference between neutral and emotional (i.e. Disgust + Sadness) evoked response, while the red line indicates the remaining significance after FDR correction. **C.** To control for the stronger deceleration observed for neutral items in the post-TNT evaluation, we computed the difference between the average heart rate of emotional and neutral pictures over the significant emotional time window extracted in the pre-TNT evaluation. The reduced or increased deceleration observed for No-Think items was still present for Disgusting and Sad pictures, respectively, even after controlling for the neutral deceleration. This suggests that the modulation of the cardiac response following memory suppression cannot be accounted by trivial attentional orientation processes. Disgust and sadness show opposite modulation. **D.** To explain the opposite modulation observed for Disgust and sadness, we split participants according to whether they could suppress or not emotional scenes from their memory. Participants who were better at suppressing Disgusting scenes from their memory also showed a reduced deceleration of cardiac response toward them afterward. On the opposite, the increased deceleration of cardiac response following the suppression of sad items was mostly related to participants who could not suppress them from their memory.

Suppression of unwanted memories inhibits later cardiac deceleration linked with disgusting scenes. We then questioned if the repeated cognitive control over emotional memories affected this typical heart rate deceleration that we observed in the pre-TNT evaluation, by comparing the heart rate changes in the post-TNT evaluation following the presentation of Think, No-Think and Baseline scenes for the three Emotional conditions (see Fig. 3 B). We predicted that suppression would reduce the cardiac response normally characterizing negative scenes. This would be reflected by less deceleration of the cardiac response following the presentation of an unpleasant scene previously suppressed and thus a positive autonomous response suppression difference compared to Baseline (No-think – Baseline time-courses). For Neutral scenes, we did not have specific predictions as to whether retrieval

suppression should further remove residual negativity (which might actually increase bpm rate following the Neutral scene), or, instead, simply dampen positive affect for the item and thus perhaps shows a more pronounced deceleration. Given the results observed in the pre-TNT emotional evaluation which shows a larger deceleration of heart rate following the presentation of negative scenes over a 1.71 to 5.32 seconds time-window, we hypothesized that the effect of memory suppression over the cardiac response should arise over a similar time-window and would induce a lesser deceleration of heart rate. This hypothesis was confirmed for Disgusting stimuli for which we found that No-Think stimuli (MDP = -1.53 bpm after 5.89 seconds) were associated with a significantly lower decrease in heart rate as compared to both Baseline (MDP = -3.16 bpm after 5.81 seconds) and Think (MDP = -3.72 bpm after 5.36 seconds) conditions. Because we expected a reduced deceleration for No-Think items, suggesting lesser emotional reaction, we compared No-Think and Baseline conditions using a one-tailed paired t-test at each time point using FDR correction for multiple comparisons. We observed a window of significance over the entire interval of presentation with the peak of maximum difference observed at 4.68 seconds ($t_{(27)} = 2.41$, P-FDR = 0.029, $d = 0.45$). This pattern of results was not observed for emotional Sadness for which No-Think items (-4.05 bpm after 4.48 seconds) augmented bpm deceleration compared to Think (-2.69 bpm after 5.33 seconds) or Baseline stimuli (-2.68 bpm after 4.16 second). However, the observed difference with baseline did not survive multiple comparisons correction ($t_{(27)} = 1.98$, P-FDR = 0.06, $d = 0.37$, one-tailed). Concerning the Neutral condition, paired t-test contrasting Think (-1.8 bpm after 5.57 seconds) and No-Think (-2.44 bpm after 5.99 seconds) with Baseline (-2.73 bpm after 5.68 seconds) did not show any significant difference (all $P_s > 0.45$, one-tailed).

The cardiac modulation associated with the neutral condition was clearly different in the post-TNT session compared to the pre-TNT session. Indeed, a deeper deceleration was now observed for neutral scenes which was clearly absent during the pre-TNT phase. This discrepancy may arise from a significant difference in cardiac frequency between these two measurements. The averaged heart rate over the periods of interest was much higher during the first evaluation (80.58 ± 10.86 bpm) than during the second one (70.80 ± 9.77 , ($t_{(27)} = -5.93$, $p < 10e-5$). This difference can be explained by the prolonged seated position during the protocol as well as a higher level of stress in the first session due to the unusual experimental situation for the participants. This diminution of about 10 bpm is enough to increase global HRV, as this measure is dependent to the initial heart rate (Mccraty & Shaffer, 2015). A higher frequency and lower HRV reduce the time between successive beats, leaving less room

for heartbeats to be modulated by attentional and stimuli processing (Pascalis, Barry, & Sparita, 1995; Abercrombie, Chambers, Greischar, & Monticelli, 2008). As a consequence, individuals with higher resting HRV often show a deeper deceleration when confronted to disgusting stimuli (Ruiz-Padial, Medialdea, del Paso, & Thayer, 2017). Here, our results are coherent with this observation as the global deceleration peak across conditions observed in the second post-TNT evaluation is more pronounced (up to -4 bpm) than in the first evaluation.

We then wanted to control for this common deceleration reflecting stimuli processing in order to isolate the specific emotional response. We subtracted the neutral deceleration from the deceleration waveform of each emotional condition and then averaged this difference over the emotional time of interest identified during the pre-TNT evaluation (see Fig. 3 C). Critically, even after controlling for this common neutral deceleration, comparisons between TNT conditions revealed that the memory control condition (No-Think) reduced the subsequent cardiac deceleration in the targeted time-window compared to Baseline ($t_{(27)} = 2.65$, $p = 0.01$, $d = 0.50$), while it induced a stronger deceleration for sad stimuli, which was also significantly lower than baseline for this emotion ($t_{(27)} = -2.21$, $p = 0.03$, $d = 0.41$). This pattern evidenced a clear interaction between memory suppression and emotional category on the subsequent modulation of cardiac activity ($t_{(27)} = 4.25$, $p < 0.001$, $d = 0.8$).

Thus, while memory suppression seems to benefit to disgusting scenes and abolishes their impact on the cardiac response (i.e. less deceleration), it paradoxically increases the cardiac deceleration observed for sad scenes. Previous work has reported that the capacity to regulate memory content and affective response are intimately related. However, as we reported in Fig 2, sad images, compared to disgusting ones, were both more difficult to control (higher intrusion rate) and less forgotten following repeated suppression attempts. We therefore hypothesized that the opposite suppression effects observed between emotional categories could in fact reflect a difference in the ability to control those images. When unsuccessful, inhibitory control may indeed worsen the emotional response associated with the targeted item (Gagnepain, Hulbert, & Anderson, 2017). It is therefore possible that the persistence of sad pictures even after repeated attempts to suppress them would be accompanied by a stronger emotional response, while the successful control of disgusting scenes could yield a parallel inhibition of both memory and cardiac activity. To assess this hypothesis, we split the participants into 2 groups according to whether or not they had forgotten the No-Think items following the TNT procedure.

For each emotion, we selected participants who had forgotten at least one out of the six items (Forgetting present group; n disgust = 16, n sadness = 11). The second group included participants who recall correctly all items in the No-Think condition (Forgetting absent; n disgust = 12, n sadness = 17) (see Fig. 3 D). The ideal procedure would have required to correct the number of forgotten items with the recall of the Baseline. However, this computation led to a small group of high suppressors in the sad condition ($n=7$) inappropriate for further analysis. Concerning emotional Disgust, the “Forgetting present” groups exhibited on average less cardiac deceleration for suppressed items compared to neutral deceleration ($t_{(15)}=1.89$, $p=0.03$, $d = 0.47$, one-tailed). This difference between the cardiac response to disgusting and neutral No-Think items was not found for the “Forgetting absent” group ($t_{(11)}=0.02$, $p=0.49$, $d = 0.006$, one-tailed). Concerning sad pictures, the greater deceleration was solely observed for the “Forgetting absent” while no difference reported for the “Forgetting present” group ($t_{(25.494)}= -2.78$, $p=0.01$, $d = 1.04$). This pattern of findings critically suggests that the ability to suppress a mental image from his memory is crucial to observe an inhibition or a reduction of the associated cardiac response. When the memory cannot be effectively suppressed as we observe for sad pictures, repeated attempts to control and exclude from awareness an unwanted memory seems, on the opposite, to have a pernicious impact and accentuate the cardiac deceleration in response to suppressed material.

Results Study 2

Study 1 evidenced that cardiac response to unpleasant images can be influenced by previous repeated inhibitory control over emotional memories. Our data suggest that this effect is specific to disgust, and that this may be due to greater mnemonic control abilities over this emotion. However, our design could nonetheless produce confounding factors explaining the modulation of cardiac activity we reported. We chose a block design for the presentation of the emotional stimuli in Study 1 (see Fig 1 A), thus grouping each Think, No-Think and Baseline condition into the same time interval. This procedure followed previous works design to contrast emotional responses (Valenza, Citi, Lanatá, Scilingo, & Barbieri, 2014). The autonomic response can be slow and long lasting, and cardiac measures can reflect either actual and preceding emotional elicitation seconds or minutes before. A grouped

presentation, plus the resting time interval between the presentation block was then suitable to control at best such contamination. However, this has the disadvantage to capture low variation linked to attentional slow process even with event-related analysis. One could for example argue that, once the participant is aware to have enter the No-Think condition, his expectation toward the stimuli differs from the other conditions and thus influence ongoing cardiac activity. To control for this possible bias, we conducted a second study adopting a semi-randomized design (see Fig. 1). We kept the separation of emotional categories in time, but randomized the presentation of Think, No-Think and Baseline stimuli inside each block. Doing so, participants were only aware of the emotional valence of the forthcoming stimuli but could not predict its condition.

Moreover, to further show that the observed effect is not linked to attentional confound occurring during emotional assessment, it is also crucial to demonstrate that the post-TNT cardiac modulation emerges as a consequence of the neural mechanism engaged during inhibitory control (i.e. TNT phase). One finding with important implications is the observation in Study 1 of substantial individual differences in the affective consequences of retrieval suppression. This observation, along with similar prior findings (Benoit, Davies, & Anderson, 2016; Gagnepain, Hulbert, & Anderson, 2017), highlights the existence of inter-individual differences in using suppression as an effective regulation and coping strategy. Thus, it would be crucial to determine if the large inter-individual differences observed in the inhibition of the cardiac response following memory suppression relate to differences in the neural mechanisms supporting inhibitory control. We then adapted the TNT procedure utilized in Study 1 to an electroencephalography (EEG) protocol. EEG has been previously used in numerous works together with the Think/No-Think paradigm, notably focusing on the oscillatory dynamic underlying memory suppression (Ketz, O'Reilly, & Curran, 2014; Depue et al., 2013; Waldhauser, Bäuml, & Hanslmayr, 2014). However, given that EEG requires a large number of trials per conditions, we limited Study 2 to disgusting pictures and did not include emotional sadness. This choice was further motivated by the difficulty to control and suppress this particular material. Twenty-four new participants underwent this procedure. The full experimental protocol is detailed in the section **Materials and methods**.

Behavioral results

Intrusions. We then tested if the item categorized as disgusting produced less intrusion than neutral scenes as we observed in Study 1. We compared the frequency of intrusion of Disgust and Neutral items over the eight TNT blocks for the No-Think condition (Fig. 4 A). An Emotion x TNT blocks ANOVA showed both an effect of Emotion ($F_{(1, 23)} = 19.49$, $\eta_g^2 = 0.01$, $p < 0.001$) and TNT blocks ($F_{(3.82, 87.88)} = 24.75$, $\eta_g^2 = 0.08$, $p < 0.001$) but no interaction between these two factors ($F_{(4.81, 110.61)} = 0.73$, $\eta_g^2 = 0.002$, $p = 0.60$).

Repeated suppression induced forgetting. An Emotion \times Condition ANOVA on participants' memories reflected by the Yes/No answer (Fig. 4 B) revealed a significant effect for Emotion ($F_{(1,23)} = 12.17$, $\eta_g^2 = 0.07$, $p = 0.002$), Condition ($F_{(1.56,35.89)} = 16.85$, $\eta_g^2 = 0.11$, $p < 0.001$) but no interaction between those two factors ($F_{(1.44,33.05)} = 2.52$, $\eta_g^2 = 0.02$, $p = 0.11$). Comparisons indicated that participants recalled significantly less No-Think item than Baseline for Disgust ($t_{(23)} = -4.40$, $p = 0.0001$, $d = 0.89$, one-tailed), while this effect was marginally significant with neutral items ($t_{(23)} = -1.59$, $p = 0.06$, $d = 0.32$, one-tailed). This result parallels previous works (Lambert, Good, & Kirk, 2010; Depue, Curran, & Banich, 2007) suggesting that emotional stimuli can be efficiently controlled and forgotten. These patterns of behavioral findings replicate observations made in Study 1 and suggest that emotional disgust may increase the motivation and the desire to reduce the momentary awareness associated with this unpleasant and unwanted memories. However, similarly to results reported in Study 1, disgusting items were globally more forgotten than neutral ones, which suggests that both a diminished memory of the items, and a stronger motivation to control them could have produced the observed patterns.

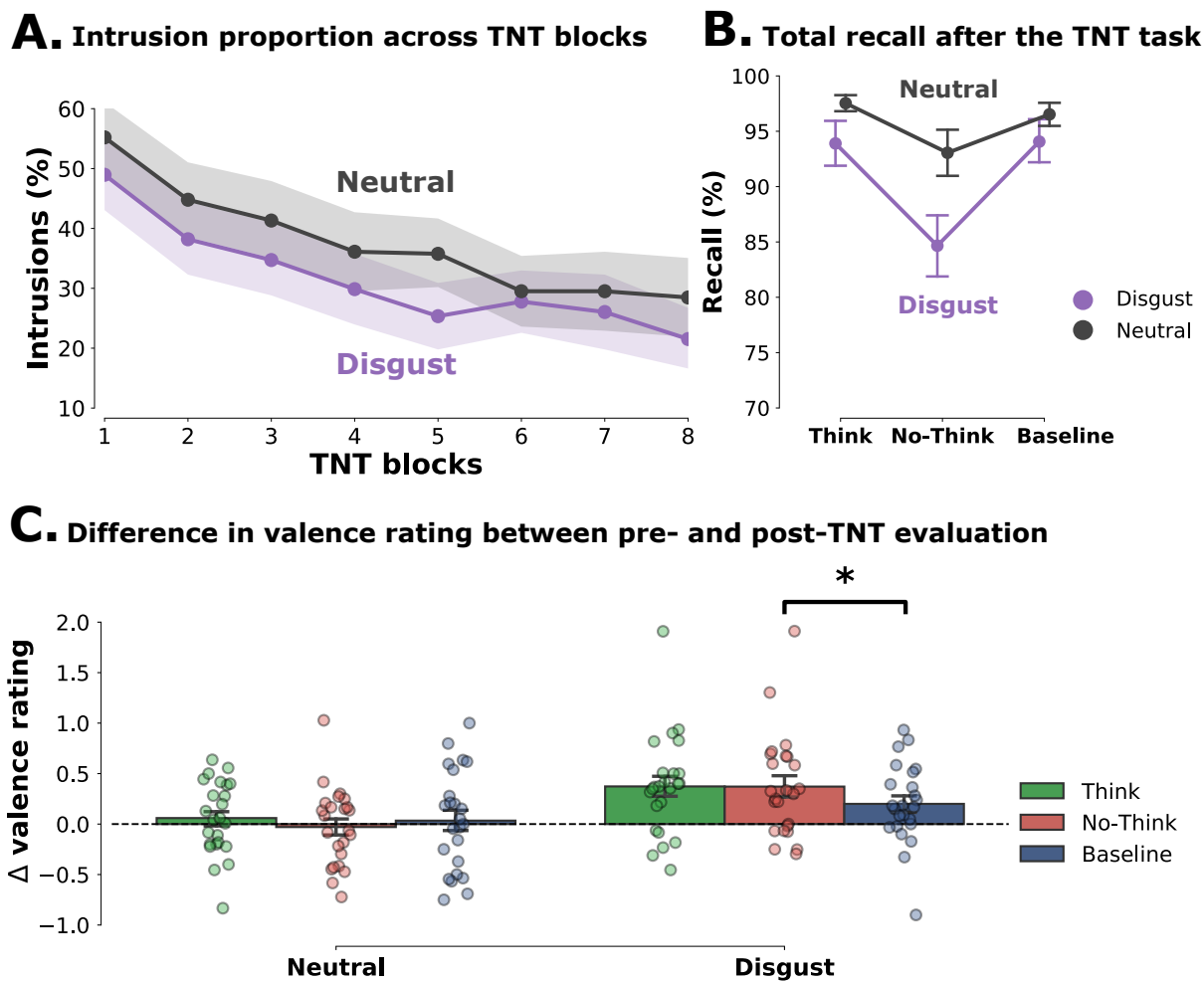


Figure 4: Behavioral indices of inhibitory control and evolution of emotional appreciation in Study 2 (n=24).

Legend Fig. 4: **A.** Intrusions proportions for No-Think trials (i.e. proportion of trials where the associated memory entered into awareness while participants were instructed to inhibit recall) over the eight TNT blocks. Participant increased their ability to control intrusion over the eight TNT blocks. On average, disgusting pictures were reported less intrusive than neutral ones. **B.** Total recall after the TNT procedure. Images in the suppression condition (No-Think) were more forgotten than baseline for disgusting items, an effect that was marginally significant for the neutral emotion. **C.** Difference in valence rating between pre- and post- TNT evaluation. We found a significant difference in valence rating between No-

Think and Baseline pictures for disgust, but this effect was not significant for neutral items. Shaded error bands and error bars represent the bootstrapped standard-error of the mean.

Effect of suppression in the subjective evaluation. We then examined the SAM valence rating after the Think/No-Think procedure adjusted by pre-TNT evaluations (Fig. 4 C). An Emotion \times Condition ANOVA revealed a significant effect of Emotion ($F_{(1, 23)} = 9.78$, $\eta_g^2 = 0.1$, $p = 0.005$) but not for Condition ($F_{(1.97, 45.38)} = 1.43$, $\eta_g^2 = 0.009$, $p = 0.25$) or the interaction between those two factors ($F_{(1.94, 44.63)} = 1.42$, $\eta_g^2 = 0.01$, $p = 0.25$). Planned comparison showed a significant higher valence rating for No-Think items as compared to Baseline items for Disgust ($t_{(23)} = 1.8$, $p = 0.04$, $d = 0.36$, one-tailed). A marginal difference was also observed for this emotion between Think, and Baseline images ($t_{(23)} = 1.66$, $p = 0.05$, $d = 0.33$, one-tailed) while no difference was found between Think and No-Think conditions ($t_{(23)} = 0.01$, $p = 0.98$, $d < 0.01$, two-tailed). Concerning Neutral scenes, we did not find similar increase of the valence rating for No-Think items as compared to baseline ($t_{(23)} = 0.74$, $p = 0.46$, $d = 0.15$). These findings show that suppressing disgusting scenes from memory may also alter the emotional quality of those memories so that their reappearance triggers less negative affect. This result parallels a recent study showing affective devaluation of words and objects after a TNT procedure (Vito & Fenske, 2017). Interestingly changes in affect do not arise for Neutral scenes suggesting that suppressing unpleasant memories may entail additional inhibitory effects not present for Neutral memories. Retrieval during Think trials measurably alters the perceived affect of the scenes to some extent, suggesting the existence of some form of habituation effect following repeated retrieval.

Physiological results

Presentation of disgusting scenes induces a decrease in heart rate. To assess the influence of an emotional elicitation on the following heart rate activity over the 10 seconds of presentation, we first compared the BPM time-course between the emotional and neutral conditions before the TNT phase (see Fig. 5 A). Replicating findings observed in Study 1, results show that presenting scenes categorized as Disgusting induced a significantly greater deceleration of the heart rate compared to Neutral pictures. Neutral scenes induced a maximal deflection peak (MDP) of -1.71 bpm 4.66 seconds after the stimuli presentation, while the deceleration was stronger for Disgusting (MDP = -2.84 bpm after 3.82 seconds) stimuli. Altogether, emotional stimuli showed a significantly larger decrease of BPM compared to

Neutral ones over a window of significance ranging from 0.81 to 4.64 seconds, with the peak of maximum difference observed at 3.28 seconds ($t_{(23)} = 3.52$, $P\text{-FDR} = 0.009$, $d = 0.71$, two-tailed paired t-test).

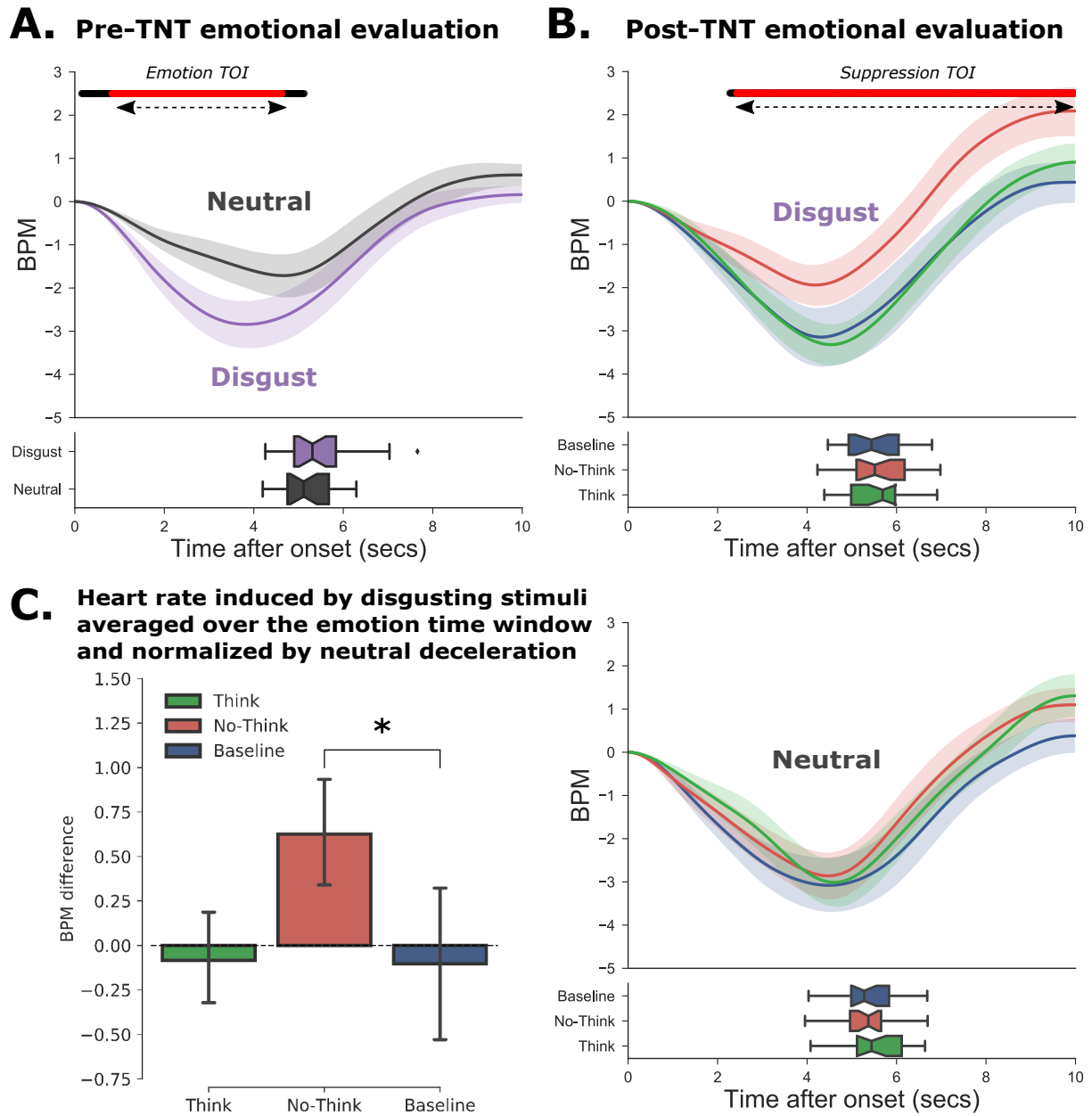


Figure 5: Cardiac response toward stimuli before and after the TNT procedure.

Legend Fig. 5: **A.** Cardiac response following the initial presentation of the images before the TNT task. The curves represent the change in the number of heartbeat per minutes (BPM) taking the onset of image presentation as a zeros starting-point. In the lower panel, the boxplots represent the distribution of response times for the valence rating. **B.** Cardiac response after the TNT task for Disgust (top panel) and Neutral (bottom panel) images. For each emotion, we represent Think, No-Think and Baseline items separately (green, red and blue lines respectively). We observed a significantly reduced deceleration of the heart rate for Disgusting No-Think items compared to Baseline. For both panel A and B, the black line shows the significant difference between neutral and disgust evoked response, while the red line indicates the remaining significance after FDR correction. **C.** To control for the stronger deceleration observed for neutral items in the post-TNT evaluation, we computed the difference between the average heart rate of disgusting and neutral pictures over the significant emotional time window extracted in the pre-TNT evaluation. The reduced or increased deceleration observed for No-Think items was still present for Disgusting pictures, respectively, even after controlling for the neutral deceleration. This suggests that the modulation of the cardiac response following memory suppression cannot be accounted by trivial attentional orientation processes.

Suppression of unwanted memories inhibits the emotionally induced cardiac deceleration. We then compared the heart rate changes in the post-TNT evaluation following the presentation of Think, No-Think and Baseline scenes for the items categorized as disgusting (see Fig. 5 B). Given the findings of Study 1, we predicted that suppression would reduce or inhibit the cardiac response normally characterizing negative scenes. This would be reflected by less deceleration of the cardiac response following the presentation of an unpleasant scene previously suppressed and thus a positive autonomous response suppression difference compared to Baseline (No-think – Baseline time-courses). Replicating findings from Study 1, we hypothesized that such a reduction of the heart autonomous response during emotional processing did occur over a time-window overlapping with the emotional effect observed during the pre-TNT emotional evaluation. For Disgusting stimuli, No-Think stimuli (MDP = -1.94 bpm after 4.18 seconds) were associated with a significantly lower decrease in heart rate as compared to both Baseline (MDP = -3.14 bpm after 4.31 seconds) and Think (MDP = -3.32 bpm after 4.53 seconds) conditions. When comparing No-Think and Baseline conditions at each time point using FDR correction for multiple comparisons, we observed a window of significance ranging from 2.41 seconds to the end of picture presentation with the peak of maximum difference observed at 7.80 seconds ($t_{(23)} = 2.45$, P-FDR

= 0.03, $d = 0.50$). No differences were found between conditions for Neutral scenes (all P s > 0.2).

Similarly to study 1, we observed an overall larger deceleration of the cardiac response to neutral images during the post-TNT evaluation compared to the pre-TNT evaluation. Again, this effect might be due to an attentional orientation effect on the cardiac response, which is more likely to occur given the lower cardiac frequency of the participants during the final post-TNT evaluation compared to the initial assessment. Indeed, the mean heart rate on the segment of interest was significantly different between the first (78.40 ± 8.72) and second evaluation (68.64 ± 10.16 , $t_{(23)} = -5.93$, $p < 10e-5$). We then controlled for the cardiac response to neutral stimuli and compared the mean heart rate inside the significant emotional interval extracted from the first evaluation (Fig. 5 C). Critically, this impact of memory suppression on the cardiac deceleration was significantly different from Baseline when controlling for neutral deceleration in the emotional time-window of interest ($t_{(23)} = 2.24$, $p = 0.03$, $d = 0.45$).

Reduction of the 5-9 Hz frequency power during cognitive control predicts cardiac modulation. We then wanted to corroborate that the observed cardiac modulation was linked to neurophysiological markers of cognitive control. We recorded the EEG signal of participants in Study 2 while they performed the TNT task. This approach has been used to investigate memory suppression in numerous studies both via analysis of event-related potential (Bergström, de Fockert, & Richardson-Klavehn, 2009; Bergström, de Fockert, & Richardson-Klavehn, 2009; Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013; Bergström, Velmans, de Fockert, & Richardson-Klavehn, 2007; Chen et al., 2012; Hanslmayr, Leipold, Pastotter, & Bauml, 2009; Hellerstedt, Johansson, & Anderson, 2016; Mecklinger, Parra, & Waldhauser, 2009) and time-frequency power-spectrum (Ketz, O'Reilly, & Curran, 2014; Depue et al., 2013; Waldhauser, Bäuml, & Hanslmayr, 2014).

We focus on the theta frequency band which is associated with memory formation (Heusser, Poeppel, Ezzyat, & Davachi, 2016; Clouter, Shapiro, & Hanslmayr, 2017) and retrieval (Nyhus & Curran, 2010; Burke et al., 2014) as well as cognitive control (Cavanagh & Frank, 2014; Mückschel, Dippel, & Beste, 2017; Eschmann, Bader, & Mecklinger, 2018). While (Depue et al., 2013), using the Think/No-Think paradigm, reported an increased theta power for

No-Think trials, (Waldhauser, Bäuml, & Hanslmayr, 2014) and (Ketz, O'Reilly, & Curran, 2014) provided divergent evidences showing a suppression of theta activity. This incongruence is thought to reflect the existence of an early and brief increase in proactive prefrontal cognitive control, and a more sustained reactive and adaptive suppression of theta oscillatory dynamics in the medial temporal lobe during the suppression trials (Waldhauser, Bäuml, & Hanslmayr, 2014). The former effect is then presupposed to occur in the early phase of the stimuli especially when the participant can anticipate up-coming reminding cue (< 1 s), while the latter is observed all along the presentation of the reminding cue.

Thus, our goal here was to quantify the inter-individual differences in controlling and suppressing the targeted memory during the TNT task using the global decrease of 5-9Hz frequency power reported by (Waldhauser, Bäuml, & Hanslmayr, 2014). We hypothesized that the mechanisms by which the suppression of memory replay might impact the later cardiac reaction should rely on this long-lasting oscillatory dynamics during the control of a memory (i.e. up to 3 s after the stimulus presentation). Indeed, this long-lasting suppression of slow oscillatory dynamics most likely reflects a desynchronization the medial temporal lobe and the nearby amygdala, whose suppression has been shown to be crucial for the inhibition of the affective response (Gagnepain, Hulbert, & Anderson, 2017). To isolate this neural marker of memory suppression, we contrasted the oscillatory power of Think and No-Think conditions associated with disgusting content. We then related this marker to the amount of cardiac inhibition observed during the subsequent post-TNT emotional evaluation. Based on similar paradigms (Ketz, O'Reilly, & Curran, 2014; Depue et al., 2013; Waldhauser, Bäuml, & Hanslmayr, 2014), we restricted our analysis to the 5-9 Hz frequency band. While theta activity is preferentially measured between 3 and 8 Hz, our goal here was to ground our analysis on previous findings to increase the reliability of subsequent comparisons and interpretations. The power of this oscillatory dynamic observed during the No-Think trials, contrasted with the values associated with memory recall (i.e. Think trials), would then reflect the neural suppression effect for each participant.

Reduction of 5-9 Hz frequency power during No-Think trials. After extracting frequency power for Think and No-Think condition of emotional images, we contrasted these conditions in order to observe the specific decrease associated with inhibitory control using non-parametric clustering statistics (Maris & Oostenveld, 2007) as implemented in MNE python. This procedure revealed 25 significant clusters among which 3 passed the

permutation statistic. The most important cluster ($p < 0.001$) lasted from 560 to 2980 ms and revealed a broad power decrease throughout No-Think as compared to Think trials (see Fig. 6 A.).

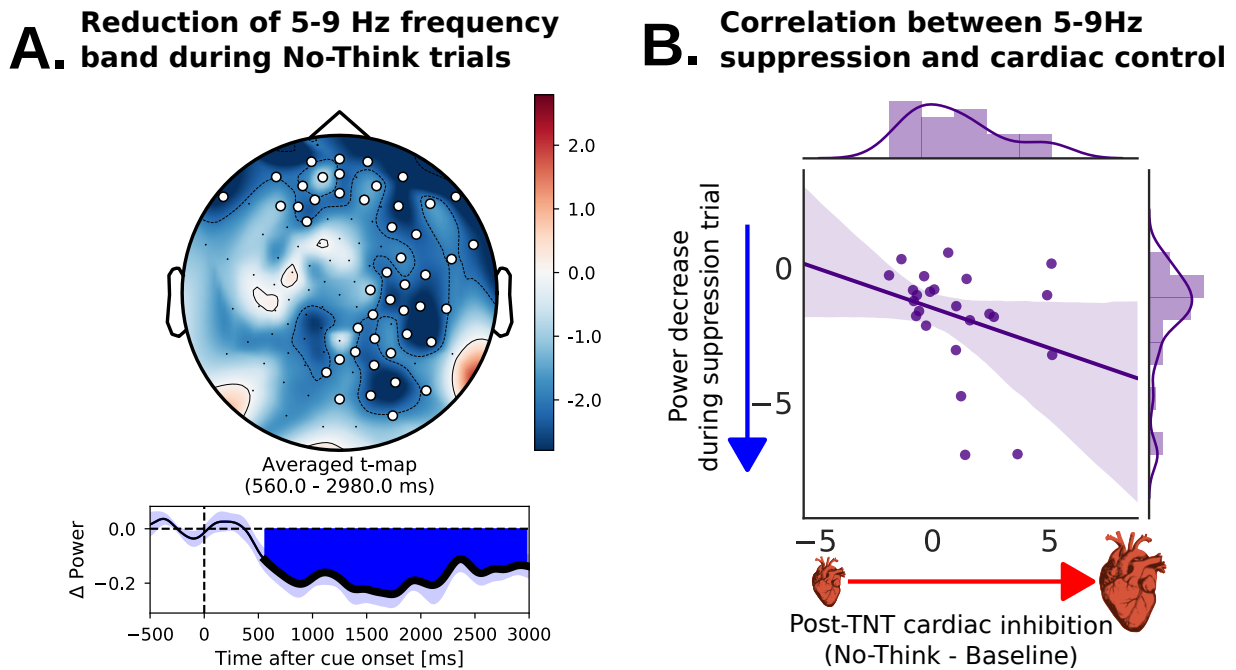


Figure 6: **Reduction of 5-9 Hz frequency power during mnemonic control and its relationship with subsequent cardiac inhibition.**

Legend Fig 6: *A. Decrease of 5-9 Hz frequency power during No-Think versus Think trials. Shaded areas represent bootstrapped standard error of the mean. Only the significant electrodes are displayed on the left panel. On the right panel, the significant time-window is represented by the bold line and the blue colored area. B. Reduction of 5-9 Hz frequency power correlates with subsequent cardiac inhibition of disgusting stimuli. This reduction is measured as the averaged difference of power between Think and No-Think trial over the significant time-window and cluster of electrodes. The cardiac inhibition is measured as the averaged heart rate difference between No-Think and baseline items over the intersection between the emotional and the suppression time-window extracted in the pre- and post-TNT evaluation, respectively.*

5-9 Hz frequency power reduction correlates with subsequent cardiac inhibi-

tion. We finally questioned if the suppression of frequency power during No-Think trials (compared to Think condition) was related to the cardiac modulation measured during the post-TNT evaluation. We performed two-tailed spearman's correlation using the robust correlation toolbox (Pernet, Wilcox, & Rousselet, 2013). This method allows to reject the null hypothesis based on the percentile bootstrap CI, an approach less sensitive to the heteroscedasticity of the data than the traditional t-test. We used skipped correlations to control for bivariate outliers. We found that the cardiac inhibition (i.e. lesser deceleration of the cardiac response to No-Think compared to Baseline scenes over the emotional time of interest identified during the pre-TNT evaluation) was negatively correlated (robust Spearman $r = -0.40$, $p < 0.05$, bootstrapped CI = [-0.72 to -0.006], no outliers detected) with the mean power difference between No-Think and Think trials (see Fig. 6 panel C). The foregoing results strongly support the hypothesis that the neural mechanism engaged to suppress retrieval doesn't simply reduce the accessibility of episodic memories but can also alter the cardiac response linked with negative affect associated with the suppressed memories. Individuals who adaptively suppressed the reactivation of unwanted memory traces could not only purge the unwanted content from awareness but also disrupt emotional content of suppressed traces at the same time, decreasing their later negative impact on the physiological response.

Discussion

Intrusive memories are preponderant in numerous psychiatric disorders and can be a source of psychological distress accompanied by autonomic manifestations of exacerbated emotional responses. The role of inhibitory control to restrain their cognitive and physiological manifestations has been debated and putatively described as worsening the emotional response or simply being ineffective. In these two studies, we showed that the successful inhibition of unwanted memories is accompanied by a long-term change of the cardiac activity associated with the presentation of the suppressed distressing material. Findings from Study 1 showed a significant inhibition of the cardiac response toward disgusting stimuli following mnemonic control and a worsening effect over the processing of sad scenes. This discrepancy was explained by greater difficulties to control sad pictures and suppress them from memory, thus supporting the relationship between forgetting and heart rate modulation. To confirm the relationship between the neural mechanisms supporting the inhibitory control of memory

awareness and subsequent cardiac inhibition, and control for potential design confound, we conducted a second study using only disgusting stimuli and during which the EEG signal was recorded to track down oscillatory dynamics supporting the inhibitory control of memories. Replicating findings from Study 2, we demonstrated a significant inhibition of cardiac activity that was specific to emotion and not observed with neutral items. Furthermore, we found a correlation between the suppression of 5-9 Hz frequency band, a prominent neural marker of memory reactivation, and the subsequent inhibition of cardiac response, thus supporting the notion that memory control and emotional regulation is supported by a common framework.

This pattern of findings consolidates previous works supporting the efficiency of direct suppression of unwanted memories. However, it extends previous works by showing that memory suppression may also impact the physiological dimension of emotions. Our results also provide a consistent argument in favor of a top-down influence of higher level of neurocognitive hierarchies on the autonomic nervous system. We interpret our findings from the perspective of the neurovisceral integration model (Thayer & Lane, 2009; Smith, Thayer, Khalsa, & Lane, 2017; Thayer & Lane, 2000) and discuss their implications for psychiatric disorders.

Cognitive control of distressing memories induces a long-term cardiac modulation. Our results show that the cardiac reaction toward disgusting memories is significantly reduced for suppressed items following the TNT task compared to baseline scenes. Alongside this effect on cardiac response, behavioral indices of both Study 1 and Study 2 show that disgusting images were successfully controlled during the TNT phase (being less intrusive than neutral or sad ones), and were subsequently more forgotten compared to baseline items (see Fig. 2 and Fig. 4). This pattern of findings suggests an influence of the memory control mechanisms over subsequent cardiac responses linked to material excluded from awareness. Study 1 used a block presentation of the Think, No-Think, and Baseline pictures during the emotional evaluation which may have potentially lead to an attentional bias for No-think items producing such an effect. In Study 2, we controlled for this effect by presenting Think, No-Think, and Baseline condition using a random order. The difference in the cardiac response to suppressed and Baseline items was still observed following TNT, suggesting that this effect is tightly coupled to the actual image presented on the screen and not due to slow variation of attentional or physiological states. This preliminary line of evidence corroborates recent studies that reported the existence of a link between the variability of the heart rate at rest and inhibitory control performances (Ottaviani et al., 2018). Such

relationship has notably been observed during the “Think/No-Think task” (Gillie, Vasey, & Thayer, 2015; Gillie, Vasey, & Thayer, 2013), thus supporting the notion that autonomic and cognitive controls are underpinned by a common framework. However, here our item-wise analyzing design allowed to relate cognitive control to the actual cardiac reaction toward the suppressed items, going further than global autonomic traits observed at rest.

Cardiac inhibition is linked to successful suppression of unwanted emotional memories. Despite these evidences, results from Study 1 also show a contrasted pattern: while cardiac inhibition was found for disgusting items, suppression of sad memories induced on the opposite a stronger deceleration. Given that we observe a stronger cardiac deceleration in response to negative scenes in the pre-TNT evaluation, this post-TNT deceleration would suggest a stronger emotional reaction which follows memory suppression. Indeed, images categorized as sad were reported as more intrusive than disgusting ones and were not more forgotten than baseline following the TNT phase (see Fig. 2). This pattern of findings suggests that direct suppression was ineffective for these stimuli. This difference cannot be explained by a lower emotional valence, based on the rating provided by either the NAPS database (Riegel et al., 2015) or by our participants. Previous works proposed that when unsuccessful, memory suppression could indeed worsen the distressing symptoms (Davies & Clark, 1998; Shipherd & Beck, 2005). To test the hypothesis that the greater deceleration we observed with sad items was due to ineffective memory suppression, we compared the cardiac modulation for sad and disgusting images of participants who effectively forgot at least one image in the post-TNT recall, with those who did not forget any (see Fig. 3). Results show that the amplitude of the cardiac modulation depended on the forgetting rate of the participants, suggesting that the worsening of the cardiac response occurred after unsuccessful control of sad memories. Thus, although cardiac modulation is influenced by memory suppression, the direction of this influence depends on the ability of the participant to successfully control and suppress distressing images from memory.

Suppression of 5-9 Hz frequency power correlates with subsequent cardiac inhibition. The current data support a tight connection between the successful control of memories and the subsequent inhibition of the cardiac response to the presentation of the suppressed scenes. However, the cardiac inhibition effect observed during the post-TNT evaluation does not necessarily imply that the memory control network has reduced some aspect of memory activation in service of affect regulation. Cardiac inhibition may instead

arise, for instance, from other sources of attentional modulation engaged during the sensory processing the suppressed scenes but independent of the neural mechanisms engaged during repeated attempts to suppress the emotional memories. Contrary to these possibilities, however, we show that the cardiac modulation correlates with the reduction of 5-9 Hz frequency band for trials where participants controlled their memory (No-Think) as compared to the condition of recall (Think), replicating earlier findings using the TNT paradigm (Waldhauser, Bäuml, & Hanslmayr, 2014). Because this oscillatory dynamic was measured with EEG in the scalp of the participants, interpreting the contribution of different cortical sources may be done cautiously. Several works show that memory encoding and retrieval is supported by an increased synchronization of slow oscillatory activity notably in the medial temporal lobe (Heusser, Poeppel, Ezzyat, & Davachi, 2016; Clouter, Shapiro, & Hanslmayr, 2017; Burke et al., 2014; Nyhus & Curran, 2010). From this perspective, the reduction of power which is associated with memory suppression might reflect an interruption and a desynchronization of memory activity. However, other neocortical sources linked to control networks may also contribute to this reduction given that similar desynchronizations in the alpha frequency band are also found when cognitive control increase (Klimesch, 2012). The decrease we report here between 5 and 9 Hz in line with (Waldhauser, Bäuml, & Hanslmayr, 2014) might indeed reflect a mixed process of both higher neocortical inhibitory engagement and a decrease of memory replay, and accurately disentangling these two components goes beyond the scope of this paper.

Implications for psychophysiology. Heart rate variability is a central manifestation of emotional experience (Kreibig, 2010). Frequency changes reflect an adaptation of cardiac activity to new environmental outcomes. According to the neurovisceral integration model (Thayer & Lane, 2009; Smith, Thayer, Khalsa, & Lane, 2017; Thayer & Lane, 2000), the amygdala, under the control of higher cortical areas (Sakaki et al., 2016; Gold, Morey, & McCarthy, 2015; Wei, Chen, & Wu, 2018; Thayer, Åhs, Fredrikson, Sollers, & Wager, 2012; Adhikari et al., 2015), plays a key role in such process (Yang et al., 2007). This structure can trigger the acceleration or deceleration of the autonomic response via direct connections with hypothalamus and brainstem circuits and a complex interplay of connections with sympathetic and parasympathetic networks (Roelofs, 2017). For example, negative experiences, which may induce hyper-activation of the amygdala, are traditionally thought to induce heart rate accelerations. This assumption hold for fear, stress or anxiety, which require a greater sympathetic recruitment to support defensive behaviors. However, several others

emotions comprising disgust or sadness, and which may also activate the amygdala, are dominated by greater parasympathetic activity, inducing a cardiac deceleration (Rohrmann, Hopp, Schienle, & Hodapp, 2009; Broderick, Grisham, & Weidemann, 2013; Shenhav & Mendes, 2014; Stark, Walter, Schienle, & Vaitl, 2005; Gilchrist, Vranceanu, Béland, Bacon, & Ditto, 2016; Meissner, Muth, & Herbert, 2011; Critchley et al., 2005). We observed this deceleratory pattern in the pre-TNT evaluation and interpreted it as a marker of emotional response toward sad and disgusting elements. Reducing cardiac activity is indeed an adaptive behavior toward various levels of threats. When facing blood-related or food-related disgusting items, this reaction could support preservation from blood loss or control of digestive reflex to prevent the ingestion of contaminated material. This is also observed with the freezing responses in many species. These reflexes are characterized by both increased sensory processing to prepare future behaviors and motor inhibition to avoid detection by predators. According to the polyvagal theory (Porges, 1995; Porges, 2001; Porges, 2011), such reactions are mediated by the parasympathetic influence of the older branch of the vagal nerve. Both Study 1 and Study 2 showed an inhibition of this deceleratory pattern for No-Think items. We assume here, in line with a previous study, that the TNT task induced a down-regulation of the amygdala for the suppressed items (Gagnepain, Hulbert, & Anderson, 2017) that is also prone to change the associated autonomic response. The inhibition of cardiac deceleration following might thus reflect a fast modulation of parasympathetic influence coherent with the recruitment of the ventral branch of the vagus (Porges, 1995; Porges, 2001; Porges, 2011).

Implications for psychiatric disorders. Our results show promising evidence for physiological regulation of disgust when memory suppression via inhibitory control is effective. The emotion of disgust has a determinant role in the development and maintenance of numerous psychiatric condition like anxiety and obsessive-compulsive disorders (Olatunji, Armstrong, & Elwood, 2017; Cisler, Olatunji, & Lohr, 2009). While this emotion does not always dominate the symptoms or induce the most salient ones, its experience can favor cognitive bias toward stimuli or situation and participates to the maintenance or increase of more dramatic manifestation like fear and state anxiety. The fear of contaminants, and the experience of disgust they induce, is for example central in obsessive-compulsive disorders (Brady, Adams, & Lohr, 2010). A high propensity to food-related disgust has also been described as a defensive mechanism to avoid caloric intake in anorexia nervosa (Vicario, 2013). These examples are straightforward as they refer to standard animal, blood or food-related stimuli, but it

is worth pointing out that disgust also encompasses a moral dimension for humans (Vicario, Rafal, Martino, & Avenanti, 2017) that we did not explore here. Moreover, while depressive syndromes do not involve a high propensity of externally-oriented disgust, self-oriented disgust is considered as a hallmark of the condition (Powell, Overton, & Simpson, 2013). Aiming at regulating this dimension could then indirectly target deeper and more resistant emotional states such as fear, sadness, anxiety traits, up to more complex moral appreciations exerting a pervasive influence on social interactions. However, although this is not clear whether more complex and resistant emotional states could benefit from training regime based on memory suppression, the current findings raise the hope that interventions focused on training mnemonic control could, in principle, reduce intrusions while dampening negative affect and physiological reaction linked to suppressed images.

By showing that retrieval suppression contributes to cardiac regulation, the present findings may offer insights into the mechanisms underlying intrusive symptoms in psychiatric disorders. Cardiac variability appears in this context as a relevant marker of both inhibitory control capabilities and emotional regulation. The direct suppression strategy, while possibly inappropriate for resistant and complex emotional states, shows promising results to regulate primal emotional experiences in the causal chain of psychiatric symptoms, and provide arguments for future behavioral and cognitive therapies where cardiac responses can inform clinical decisions.

Conclusion

In this work, we explored the influence of inhibitory control on the physiological reaction associated with emotional memories. Replicating a converging set of evidence showing an increased cardiac deceleration in reaction to the presentation of images classified as disgusting or sad, we show that this pattern was inhibited after direct suppression using the Think/No-Think protocol. This effect was linked to the ability of the participant to control the intrusion of the memories during the task as well as the reduction of specific oscillatory dynamic during suppression, suggesting a lesser reactivation of the memories. Taken together, our findings support the notion that inhibitory control engages targeted regulation of emotional memories both in the experiential level and evidence for the first time its influence over the autonomic embodiment of emotions. Given the prevalence of distressing intrusion, notably mediated by reaction of disgust, in the origin and maintenance of numerous psychiatric disorders,

this work opens avenues for the development of markers and cognitive training integrating memory control and psychophysiological measures.

Materials and Methods

Study 1

Participants. Twenty-eight right-handed native French speakers between the ages of 18 and 35 were paid to participate (14 males). They had no reported history of neurological, medical, visual, or memory disorders. The study was approved by the local Ethics Committee (CPP Nord Ouest I, N° ID RCB: 2015-A01727), and all participants gave written consent. Participants were asked not to consume psychostimulants, drugs, alcohol or coffee during the hours preceding the experimental period and to avoid intense physical activity the day before.

Material. The stimuli were 72 object-scene pairs plus six fillers pairs for think/no-think practice. The objects were selected from the Bank of Standardized Stimuli (BOSS) (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014) and depicted only nonliving and non-organic artifacts. The scenes were selected from the Nencki Affective Picture System (NAPS) data base (Marchewka, Łukasz Żurawski, Jednoróg, & Grabowska, 2013; Riegel et al., 2015; Wierzba et al., 2015). Half of the critical scenes were normed as Neutral (36 images, mean valence: 5.79; SD valence: 0.84; mean arousal: 4.78; SD arousal: 0.71). The other half were normed as Negative, and induced either a feeling of Disgust (18 images, mean valence: 3.02; SD valence: 0.76; mean arousal: 6.42; SD arousal: 0.58) or Sadness (18 images, mean valence: 2.96; SD valence: 0.49; mean arousal: 6.43; SD arousal: 0.64). We referred to the discrete quotation provided in the NAPS data base (Wierzba et al., 2015) to select the images. However, as we aimed at obtaining an homogeneous set where each scene depicted contents as unique as possible, we manually completed this selection with non-normed image from the NAPS. The Neutral and Negative lists were matched in terms of number of Animals, Faces, Objects and People. Each object was carefully paired with a scene to avoid the existence of any pre-existing semantic association, which could have biased memory encoding. Three lists of 12 pairs (assigned to Think, No-Think, and Baseline conditions) were created for each valence condition and were counterbalanced so that they appeared in each TNT condition equally often, across participants.

Pre-TNT emotion evaluation. To measure the initial emotional reaction to selected scenes and provide a better estimation of the impact of subsequent TNT manipulation on affective responses, participants first performed an emotional rating task (Figure XX). This task consisted in rating both the valence and arousal associated with each visual scene using the Self-Assessment-Manikin (SAM) pictorial scale (Lang, 2008). These 9-point scales quantify participants' self-reported valence and arousal response to each visual scene reported on a scale labeled with diagram-like manikins with varying emotional facial expressions.

Each scene was presented on grey background for a total of 10 seconds followed by a 500 ms inter-stimuli interval. During the first 3 seconds, the scene was presented without any scale and participants were instructed to carefully explore the scene and understand its meaning. Following these initial 3 seconds, both the valence and arousal SAM pictorial scales appeared underneath the picture for 7 seconds. The first scale right underneath the picture corresponded to the valence scale. This scale ranges from 1 (corresponding to a frowning face on the far left of the scale) if a picture made them feel completely unhappy to 9 (corresponding to a smiley face on the far right if a picture made them feel completely happy). If participants felt neutral, neither happy nor sad, they were then instructed to press the square under the figure in the middle. The arousal scale was presented at the bottom of the screen and ranges from 1 (corresponding to a calm face on the far left of the scale) if a picture made them feel completely relaxed to 9 (corresponding to an excited face on the far right if a picture made them feel completely aroused). If participants felt neutral, neither relaxed nor aroused, they were then instructed to press the square under the figure in the middle position. Participants selected using the computer mouse the facial expression most closely matching their perception of the valence and the arousal of the scene. Participants were instructed to rate the valence first because we were primarily interested by the impact of subsequent TNT manipulation on the adverseness of the suppressed memories.

We recorded the electrocardiogram (ECG) using the Biopac MP36R Data Acquisition system and its PC software AcqKnowledge 4.1 during the full duration of this task (~20 min.). Although the presentation of these data goes beyond the scope of this paper, the electrodermal activity (EDA), were also recorded. All the participants undertook this emotional rating task in the morning (8 a.m) and were familiarized with the task settings using 10

independent images selected in the NAPS database. After an initial resting block of 2 min., the 72 images used in the think/no-think procedure (including baseline items) were presented according to two blocks of Neutral and Negative scenes whose order was counterbalanced across participants. The presentation of these two main blocks was further separated by a resting block of 2 min. designed to prevent potential overlap and contamination between the cardiac responses to Negative and Neutral scenes. Each block of Neutral and Negative scenes was further divided into three sub-blocks of 12 images (126 seconds per sub-block) corresponding to Think, No-Think, and Baseline condition. Each of these sub-blocks was followed by a resting period of 30 sec and the order of assignment to TNT condition was counterbalanced across participants (see Fig 1).

Learning and test/feedback. After this procedure, participants learned the 72 object-scenes pairs used in the think/no-think task plus the six training pairs. Participants first learned all object-scene pairs through a test-feedback cycle procedure. After studying all pairs for 6 s each, participants were given test trials presenting the object cue for a given pair for up to 4 s and asked whether they could recall and fully visualize the paired scene. If so, three scenes then appeared (one correct and two foils were randomly taken from other pairs of the same emotional valence), and they received up to 4 s to select which scene went with the object cue. After selecting a scene or if the response window expired, a screen appeared for 1 s indicating whether the recognition judgment was correct or incorrect. In all cases (even if participants indicated that they could fully visualize the associated scene in the first step), each trial ended with the correct pairing appearing onscreen for 2.5 s. Participants were asked to use this feedback to increase their knowledge of the pair. Once all pairs had been tested, the percentage of total recall was displayed on the screen. If this percentage was inferior to 95% in at least one emotional condition, all the pairs were presented again in a randomized order. This procedure was repeated until the score was superior to 95%, thus ensuring correct encoding of the pairs and a comparable exposure to all the items. This overtraining procedure was intended to ensure that images would intrude when the cue was presented during the think/no-think task and further reduces any encoding advantage negatively valenced scenes may have.

Criterion test. Following learning and after practicing with the think/no-think task, participants performed an attentional task for 20 minutes using an entirely independent image set and aiming to address a different question (data not reported here). At this point, parti-

Participants viewed a brief reminder of all of the studied pairs (3 s each), during which they were asked once again to reinforce their knowledge of the pairings. Following this pair refresher, a final criterion test was proposed where the participant had to recall the correct image in a similar way to the test/feedback procedure, but this time without feedback and only once. This allowed to isolate items forgotten before the think/no-think task and to exclude them from the analysis.

Think/no-think task. Participants then performed the think/no-think task, which was divided into 8 TNT blocks, each 10–11 min in length. Each block consisted in the presentations of the 24 “think” and 24 “no-think” items (12 Neutral, 6 Disgust and 6 Sadness each), yielding, across the 8 blocks, 256 trials for Neutral and 256 for Negative conditions (128 Disgust + 128 Sadness). Cues appeared for 3 s either framed in green or red, centered on a grey background. On “think” trials, the cue was bounded by a green box, and participants were told to generate an image of the associated scene as detailed and complete as possible. On “no-think” trials, the cue was bounded by a red box, and participants were told that it was imperative to prevent the scene from coming to mind at all and that they should fixate and concentrate on the object-cue without looking away. During red-cued trials, participants were asked to block thoughts of the scene by blanking their mind and not by replacing the scene 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 offset of each of the Think or No-Think trial cues, participants reported if the associated scene had entered awareness by pressing the button under their right index if it did, or the button under their right middle finger if it did not. Participants had up to 2 seconds to make this rating, so they were instructed and trained to make this quickly without thinking about the associated picture. Their response was followed by a jittered fixation cross randomly lasting between 2400 to 3600 ms. These “intrusion ratings” were used to isolate trials with intrusive memories and quantify their occurrence. For each repetition of a given condition (e.g., No-Think/Disgust), we averaged these binary intrusion reports across all items in that condition to compute an intrusion proportion for that repetition. We then averaged these intrusion proportions across the 8 TNT blocks to derive the overall intrusion rate for a given participant.

Recall. After this procedure, after effects of memory suppression were examined via a cued-recall task on all the object-scene pairs, including the 24 Baseline scenes (12 Neutrals and 12 Negatives) omitted from the think/no-think task. Because these items were trained at

the same time as the Think and No-Think items but did not participate in the think/no-think task, they provide a baseline estimate of memorability of the scenes, given that they are similarly old but neither suppression nor retrieval had been performed on them. During trials of this cued-recall task, the cue object was presented in the center of the screen for 5 seconds and participant were told to visually recall the associated picture. If they could recall the associated scene, they were told to press the button under their right index finger and then had 15s to verbally describe the scene in as much detail as possible. The descriptions were recorded. If they could not recall the associated scene, they had to press the button under their right middle finger, and the next object was presented to them. Participants' descriptions were scored checked to ensure enough coherence between the described item and the real pictures. On average, participants made false recall for 0.3% of the images in Study 1 and for 0.8% in Study 2.

Post-TNT emotion evaluation. Following the recall task, participants were asked to rate the 72 pictures using the same procedure to measure whether retrieval suppression influenced later affective responses to scenes. During this task, all of the scenes (“Think”, “No-think” and “Baseline”) were presented again and we asked participants to give their perceived emotional valence and arousal while physiological signals were recorded. This procedure enables us to determine whether retrieval or suppression altered the valence, arousal or the cardiac response associated with the item, relative to “Baseline” pairs. Because the SAM valence scale includes “neutral” ratings at mid-scale [with negative affect on the low end (score 1–4), and positive affect on the high end (scores 6 –9)], higher valence scores for suppressed Negative scenes likely reflect diminished negative affect (e.g., moving from a 1 to a 3 involves a movement toward neutral). In contrast, higher valence scores for neutral scenes likely reflect increasing positive affect. We used the foregoing ratings to compute valence scores for each condition for each participant, adjusted for pre-TNT valence. This adjustment assumes that the SAM valence rating that a participant gives for a scene is likely to be influenced by three contributing components: the scene’s initial valence, the effects of our experimental manipulation for that participant (here, the effects of the think/no-think task) and noise (e.g., attentional fluctuation).

Study 2

We recruited twenty-four participants (12 males) according to the same criterion than Study 1. We conducted Study 2 in order to replicate the main results observed in Study 1 (heart rate modulation induced by inhibition of unwanted memories), control for possible experimental confounds (see the description of Study 2 below and the Results section). We measured the EEG response during the TNT task in order to estimate the relationship between the individual neural markers of mnemonic control and the aftereffects of memory suppression on the cardiac response. The differences with Study 1 are detailed in Figure 1.

Material. Study 1 initially included 18 scenes for each emotion, which were divided across Think, No-Think, and Baseline conditions. To increase the number of scenes associated with each condition of interest, we restricted our material in Study 2 to 36 Neutral and 36 Disgust stimuli plus 6 more filler pairs for practice. The stimuli were selected from the same database than Study 1 with half of the critical scenes rated as Neutral (mean valence: 5.79; SD valence: 0.84; mean arousal: 4.78; SD arousal: 0.71) and the other half as Disgust (mean valence: 3.15; SD valence: 0.68; mean arousal: 6.35; SD arousal: 0.48).

Pre- and Post-TNT emotional evaluations. Emotional assessment in Study 1 included a block design during which stimulus presentation alternate between discrete epochs of scenes drawn from the same condition (i.e. Think, No-Think, and Baseline). This design was intended to optimize the detection of differences in heartbeats which may have only arisen in the low-frequency range. One potential problem with blocked designs, however, is that the response to events within a block may be confounded by the context of their occurrence (e.g. when participants become aware of the blocking and may alter their strategies/attention as a consequence). To control for that potential bias, Study 2 used a design in which the presentation of Think, No-Think, and Baseline items was randomized and unpredictable. However, we kept emotion and neutral conditions presented in separated blocks to control for long-term autonomic change induced by emotions that may spread over and modulate cardiac response to neutral scenes. In addition, we also restricted the subjective assessment of emotional response to the valence rating task (and removed the arousal scale). This last change explains the difference in mean response time between Study 2 and Study 1.

Think/no-think task. Electroencephalographic (EEG) activity was recorded during the think/no-think task of Study 2 (data not reported here). Application of the EEG cap (Electrical Geodesics, INC., Hydrocel Geodesic Sensor Net) was fast (about 15mn) causing a very trivial delay between learning and TNT task. After the offset of each of the Think or No-Think trial cues, participants reported the extent to which the associated scene had entered awareness by using a 3-point scale, instead of binary yes/no choice in Study 1, corresponding to the labels: never, briefly, often. Although we used participants' responses to classify each No-Think cue as either an intrusion (i.e. a "briefly" or "frequent" response) or not (a "never" response) in a binary fashion, this 3-point scale was proposed to increase participants' awareness toward intrusive memory and ensure that participants would engage appropriate inhibitory resources throughout the task to countermand reminiscence of memory whether the very briefly, or strongly, pops into participant's mind. In addition, while participants had no explicit time constraint to give their feedback during this procedure, we asked them to make these reports as quickly and intuitively as possible. We do so to avoid that they dwell too much on the intrusion and finally recall the memories.

EEG

Procedure

All the participants performed the TNT task inside a Faraday's cage. Stimulus presentation was controlled by Eprime Pro on a 17" LCD screen with a 1280 x 1024 resolution. To avoid exploratory eye movement, each picture was displayed inside a 400 x 400 pixel square. The square contour was colored in green or in red and represented the "Think" and "No-Think" condition respectively. Participants were seated comfortably in a dimly lit room during the whole experiment at a distance from 90 cm to 100 cm from the screen and were instructed to minimize blinking and moving during recording.

The cue picture was presented on a black background during 3000ms and was followed by a white fixation cross. Participant had to give a feedback concerning the intrusion rate on a response box using their right hand. Three responses were possible: **1**-No intrusion at all, **2**-The image intruded only for a short amount of time and/or was successfully controlled, **3**-The image intruded vividly for a large amount of time. The fixation cross stayed on the screen

until the participant's feedback. This phase lasted 42.01 ± 3.96 minutes (mean \pm standard deviation) and was divided into 8 TNT blocks. After two consecutive blocks, a message was displayed in a black screen indicating to the participant that he/she could rest. The participant decided when he/she wanted to resume the experiment by pressing any button on the response box. After the first phase, the experimenter entered in the Faraday's cage with the participant to propose a last questionnaire concerning the performances on the TNT task and ensure that the participant is performing it correctly.

EEG recording

EEG activity was recorded continuously by GES 300 amplifier (Electrical Geodesics, INC.) using an EGI Hydrocel Geodesic Sensor Net (HGSN-128) with a dense array of 128 Ag/AgCl sensors (Tucker, 1993). Impedances were kept under 100 k (Ferree, Luu, Russell, & Tucker, 2001), and EEG channel was referenced to a vertex reference Cz and ground to CPPZ (fixed by the EGI system). The signal was sampled at 20 kHz frequency with a 24-bit A/D and was online (hardware) amplified and low pass filtered at 4 KHz. The signal was filtered by a Butterworth low pass Finite Impulse Response (FIR) filter at 500 Hz and sub-sampled at 1 KHz. Electro-oculogram was recorded using four electrodes placed vertically and horizontally around the eyes. EEG data were processed offline using Netstation 4.4.2 (Electrical Geodesics Inc., Eugene, OR, USA). The signal was filtered using a 1 Hz Kaiser FIR first-order high-pass filter (which ensures a linear phase and no distortion in the bandwidth) in order to discard DC and very slow waves.

EEG analysis

Preprocessing. EEG analyses were performed using the version 0.15 of the MNE library (Gramfort, 2013; Gramfort et al., 2014) implemented in python 3.6 and followed recent recommended analysis steps (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015; Andersen, 2018; Jas et al., 2018). With the exception of bad channel removal and EOG selection, which was performed by visual inspection, the EEG pipeline relies on automated steps in order to improve both sharing of the results and further replication. We removed 13 peripheral electrodes (F11, FT11, T9, TP11, P09, I1, I2, P010, TP12, T10, FT12, F12), out of the 115 EEG electrodes, due to global bad impedance quality or no contact with the scalp at all. We low-pass filtered the raw data at 40 Hz cutoff frequency using a

Finite Impulse Response (FIR) filter and referenced to a common average. On the remaining 102 channels, sensors showing no signal, constant white noise or intermittent signal, as observed during the EEG recording, were interpolated. Cue onset was adjusted for the delay (+15ms) between trigger generation and image appearance on the screen in the Faraday's cage, measured using photodiodes during preliminary tests. We then segmented the trials from -1000ms to 3500ms around the cue onset. We systematically rejected trials with a large threshold of 350 μ V before detecting and correcting artifacts using the Autoreject library (Jas, Engemann, Bekhti, Raimondo, & Gramfort, 2017) with local thresholds detection and the Bayesian optimization method. Finally, we applied ICA correction to reduce the remaining eye blink artifacts based on one manually selected EOG channel.

Time-frequency. Our analysis focused on the contrast between No-Think and Think conditions during the TNT task for items classified as disgusting. For the 'Think' condition, we selected only trials where the participant reported a successful memory reactivation (feedback = 1). We kept all available 'No-Think' trials and extracted the power for each epoch in the high theta frequency band [5-9 Hz] in order to replicate results from (Waldhauser, Bäuml, & Hanslmayr, 2014). To do so we similarly applied 5 Morlet wavelets in frequencies ranging from 5 to 9 Hz with a width of 5 cycles. After this procedure, we cropped the epochs from -0.5 s to 3.0 s to avoid edge artifacts and baseline corrected by subtracting the mean of the response in the interval from -500 to 0 ms. The resulting time-frequency data were then down-sampled to 10 ms time-bins for further analysis.

Statistical analysis

Analyses were performed using custom Python scripts (Python 3.6) together with specific modules for EEG analysis (Gramfort, 2013) and R packages (R version 3.4) for factorial analysis (Singmann, Bolker, Westfall, & Aust, 2016) and effect size calculation (Torchiano, 2017). Robust correlations were performed using the correlation toolbox (Pernet, Wilcox, & Rousselet, 2013) implemented in Matlab 2017b (The Mathworks). Figures, scripts, statistical analysis and the data reported in this paper are made available at: https://github.com/legrandnico/cardiac_contr

Behavior. Items that were not correctly recalled during the final test before TNT were systematically excluded from later analyses (Study 1: 1.2%; Study 2: 1.1%). We ana-

lyzed behavioral results with ANOVA integrating all the available factors and tested specific hypotheses with planned comparisons. We performed two-sided comparisons when no *a priori* hypothesis was formulated and one-tailed when we expected specific effects. In order to improve the clarity of our presentation, figures do not report all the possible comparisons but only relevant effects.

ECG. Comparisons between Neutral and Emotional heart rate curves were made using a paired t-test at each time point, we used the *fdr_correction()* function implemented in MNE Python (Gramfort, 2013) to correct for multiple comparisons. For each group, we set the significance threshold at 0.05 to assess the effect of emotions in the first session and used a one-tailed t-test for measuring the effect of suppression in session 2.

EEG. We subtracted the mean values observed at each time point and for each electrode under the ‘Think’ condition to the ‘No-Think’ condition. We then performed a non-parametric spatiotemporal cluster statistic (Maris & Oostenveld, 2007) using implementations into MNE python as described in (Jas et al., 2018). We used a two-tailed t-test as the statistic on which to perform the clustering with permutation (n=1024). As multiple clusters were outputted as significant, we automatically selected the cluster with the longest time duration.

Acknowledgements

We thank Lea Lemoine for his assistance with data preprocessing in Study 1, Emmanuelle Duprey and Camille Rebillard for the medical examinations and Florence Fraisse for administrative support. This work was supported by a grant from the Observatoire B2V des memoires for N.L.

Author contribution

N.L. programmed the experiment, collected and analyzed the data, and wrote the paper. P.G. designed the experiment, supervised data analysis, and wrote the paper. O.E. designed the experiment, supervised data collection, and provided inputs for data analysis and

manuscript redaction. M.P., A.V., and F.V. performed medical inclusion visit. P.C. and F.D. supervised and performed EEG data collection. D.P. and F.E. supervised N.L.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

Images

Study 1

NEUTRAL NAPS reference: Animals [_043.h, _088.h, _097.h, _105.h, _108.h, _132.h, _155.v, _213.h], Faces [_060.h, _061.h, _065.h, _123.h, _184.h, _211.h, _219.v, _292.h, _306.v], Objects [_029.h, _049.h, _164.h, _200.h, _242.h, _270.h], People [_026.h, _028.h, _035.h, _091.h, _095.h, _097.h, _100.h, _148.h, _149.h, _151.h, _164.h, _179.h, _249.h].

DISGUST NAPS reference: Animals [_039.h, _041.h, _062.h, _065.h], Faces [_156.h], Objects [_006.h, _010.h, _011.h, _022.h, _125.h, _126.h], People [_019.v, _077.v, _090.v, _205.v, _209.v, _220.h, _239.h].

SADNESS NAPS reference: Animals [_013.h, _053.h, _067.h, _077.h], Faces [_011.h, _012.v, _013.h, _019.h, _034.h, _149.v, _279.h, _294.h], People [_001.h, _002.v, _127.h, _133.h, _143.h, _147.h].

Study 2

NEUTRAL NAPS reference: Animals [_043_h, _088_h, _097_h, _105_h, _108_h, _132_h, _155_v, _213_h], Faces [_060_h, _061_h, _065_h, _123_h, _184_h, _211_h, _219_v, _292_h, _306_v], Objects [_029_h, _049_h, _164_h, _200_h, _242_h, _270_h], People [_026_h, _028_h, _035_h, _091_h, _095_h, _097_h, _100_h, _148_h, _149_h, _151_h, _164_h, _179_h, _249_h].

DISGUST NAPS reference: Animals [_008_v, _018_h, _039_h, _041_h, _062_h, _065_h, _078_h], Faces [_156_h, _296_h, _368_h], Objects [_006_h, _007_h, _010_h, _011_h, _013_h, _021_v, _022_h, _023_v, _088_h, _106_v, _125_h, _126_h, _154_h], People [_019_v, _077_v, _090_v, _204_v, _205_v, _209_v, _213_v, _217_h, _220_h, _223_h, _230_h, _239_h, _247_v].

References

- 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. <http://doi.org/10.1037/a0018113>
- Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, *26*(1), 1–26. <http://doi.org/10.1080/1047840x.2014.940781>
- Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience*, *16*(11), 693–700. <http://doi.org/10.1038/nrn4044>
- Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. <http://doi.org/10.1016/j.cpr.2009.11.004>
- Fernandez, K. C., Jazaieri, H., & Gross, J. J. (2016). Emotion Regulation: A Transdiagnostic Perspective on a New RDoC Domain.. *Cognit Ther Res*, *40*, 426–440.
- Berking, M., & Wupperman, P. (2012). Emotion regulation and mental health: recent findings, current challenges, and future directions.. *Curr Opin Psychiatry*, *25*, 128–34.
- Shepherd, L., & Wild, J. (2014). Emotion regulation, physiological arousal and PTSD symptoms in trauma-exposed individuals.. *J Behav Ther Exp Psychiatry*, *45*, 360–7.
- Fergus, T. A., & Bardeen, J. R. (2014). Emotion regulation and obsessive–compulsive symptoms: A further examination of associations. *Journal of Obsessive-Compulsive and Related Disorders*, *3*(3), 243–248. <http://doi.org/10.1016/j.jocrd.2014.06.001>
- Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: relation to cognitive inhibition.. *Cogn Emot*, *24*, 281–98.
- Mihailova, S., & Jobson, L. (2018). Association between intrusive negative autobiographical memories and depression: A meta-analytic investigation. *Clinical Psychology & Psychotherapy*. <http://doi.org/10.1002/cpp.2184>
- Wegner, D. M., & Zanakos, S. (1994). Chronic Thought Suppression. *Journal of Personality*, *62*(4), 615–640. <http://doi.org/10.1111/j.1467-6494.1994.tb00311.x>
- Dunn, B. D., Billotti, D., Murphy, V., & Dalgleish, T. (2009). The consequences of effortful emotion regulation when processing distressing material: A comparison of suppression and

- acceptance. *Behaviour Research and Therapy*, 47(9), 761–773. <http://doi.org/10.1016/j.brat.2009.05.007>
- Moritz, S., Hörmann, C. C., Schröder, J., Berger, T., Jacob, G. A., Meyer, B., ... Klein, J. P. (2014). Beyond words: sensory properties of depressive thoughts.. *Cogn Emot*, 28, 1047–56.
- Rachman, S. (2007). Unwanted intrusive images in obsessive compulsive disorders.. *J Behav Ther Exp Psychiatry*, 38, 402–10.
- Ólafsson, R. P., Snorrason, Í., Bjarnason, R. K., Emmelkamp, P. M., Ólason, D. ., & Kristjánsson, Á. (2014). Replacing intrusive thoughts: investigating thought control in relation to OCD symptoms.. *J Behav Ther Exp Psychiatry*, 45, 506–15.
- Campbell-Sills, L., Barlow, D. H., Brown, T. A., & Hofmann, S. G. (2006). Effects of suppression and acceptance on emotional responses of individuals with anxiety and mood disorders. *Behaviour Research and Therapy*, 44(9), 1251–1263. <http://doi.org/10.1016/j.brat.2005.10.001>
- Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology.. *J Pers Soc Psychol*, 74, 224–37.
- Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., & Banich, M. T. (2016). The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive, Emotional, and Motor Processes.. *Cereb Cortex*, 26, 1634–1646.
- Cole, M. W., Repovš, G., & Anticevic, A. (2014). The Frontoparietal Control System. *The Neuroscientist*, 20(6), 652–664. <http://doi.org/10.1177/1073858414525995>
- Anderson, M. C. (2004). Neural Systems Underlying the Suppression of Unwanted Memories. *Science*, 303(5655), 232–235. <http://doi.org/10.1126/science.1089504>
- Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process. *Science*, 317(5835), 215–219. <http://doi.org/10.1126/science.1139560>
- Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences*, 111(13), E1310–E1319. <http://doi.org/10.1073/pnas.1311468111>

- Benoit, R. G., & Anderson, M. C. (2012). Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories. *Neuron*, *76*(2), 450–460. <http://doi.org/10.1016/j.neuron.2012.07.025>
- Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive Top-Down Suppression of Hippocampal Activity and the Purging of Intrusive Memories from Consciousness. *Journal of Cognitive Neuroscience*, *27*(1), 96–111. http://doi.org/10.1162/jocn_a_00696
- Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., & Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Human Brain Mapping*, *28*(5), 409–423. <http://doi.org/10.1002/hbm.20291>
- Hayes, J. P., Morey, R. A., Petty, C. M., Seth, S., Smoski, M. J., McCarthy, G., & LaBar, K. S. (2010). Staying Cool when Things Get Hot: Emotion Regulation Modulates Neural Mechanisms of Memory Encoding. *Frontiers in Human Neuroscience*, *4* (December), 1–10. <http://doi.org/10.3389/fnhum.2010.00230>
- Kim, S. H., & Hamann, S. (2007). Neural Correlates of Positive and Negative Emotion Regulation. *Journal of Cognitive Neuroscience*, *19*(5), 776–798. <http://doi.org/10.1162/jocn.2007.19.5.776>
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D. E., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *NeuroImage*, *23*(2), 483–499. <http://doi.org/10.1016/j.neuroimage.2004.06.030>
- Kohn, N., Eickhoff, S. B., Scheller, M., Laird, A. R., Fox, P. T., & Habel, U. (2014). Neural network of cognitive emotion regulation — An ALE meta-analysis and MACM analysis. *NeuroImage*, *87*, 345–355. <http://doi.org/10.1016/j.neuroimage.2013.11.001>
- Comte, M., Schön, D., Coull, J. T., Reynaud, E., Khalfa, S., Belzeaux, R., . . . Fakra, E. (2016). Dissociating Bottom-Up and Top-Down Mechanisms in the Cortico-Limbic System during Emotion Processing. *Cerebral Cortex*, *26*(1), 144–155. <http://doi.org/10.1093/cercor/bhu185>
- Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., & Banich, M. T. (2015). The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive Emotional, and Motor Processes. *Cerebral Cortex*, *26*(4), 1634–1646.

<http://doi.org/10.1093/cercor/bhu324>

Depue, B. E., Burgess, G. C., Willcutt, E. G., Ruzic, L., & Banich, M. T. (2010). Inhibitory control of memory retrieval and motor processing associated with the right lateral prefrontal cortex: evidence from deficits in individuals with ADHD.. *Neuropsychologia*, *48*, 3909–17.

Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories. *The Journal of Neuroscience*, *37*(27), 6423–6441. <http://doi.org/10.1523/jneurosci.2732-16.2017>

James, W. (1884). What is an emotion?. *Mind*, *9*(34), 188–205. <http://doi.org/10.1093/mind/os-ix.34.188>

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, *14*(2), 143–152. <http://doi.org/10.1038/nrn3403>

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655–666. <http://doi.org/10.1038/nrn894>

Garfinkel, S. N., & Critchley, H. D. (2016). Threat and the Body: How the Heart Supports Fear Processing. *Trends in Cognitive Sciences*, *20*(1), 34–46. <http://doi.org/10.1016/j.tics.2015.10.005>

Babo-Rebelo, M., Richter, C. G., & Tallon-Baudry, C. (2016). Neural Responses to Heartbeats in the Default Network Encode the Self in Spontaneous Thoughts. *The Journal of Neuroscience*, *36*(30), 7829–7840. <http://doi.org/10.1523/jneurosci.0262-16.2016>

Pezzulo, G., Iodice, P., Barca, L., Chausse, P., Monceau, S., & Mermillod, M. (2018). Increased heart rate after exercise facilitates the processing of fearful but not disgusted faces. *Scientific Reports*, *8*(1). <http://doi.org/10.1038/s41598-017-18761-5>

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. <http://doi.org/10.1016/j.biopsycho.2010.03.010>

Ekman, P., Levenson, R., & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210. <http://doi.org/10.1126/science.6612338>

Valenza, G., Citi, L., Lanatá, A., Scilingo, E. P., & Barbieri, R. (2014). Revealing Real-Time

- Emotional Responses: a Personalized Assessment based on Heartbeat Dynamics. *Scientific Reports*, 4(1). <http://doi.org/10.1038/srep04998>
- Critchley, H. D., Rotshtein, P., Nagai, Y., O'Doherty, J., Mathias, C. J., & Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *NeuroImage*, 24(3), 751–762. <http://doi.org/10.1016/j.neuroimage.2004.10.013>
- Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410(6826), 366–369. <http://doi.org/10.1038/35066572>
- Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75, 274–296. <http://doi.org/10.1016/j.neubiorev.2017.02.003>
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, 61(3), 201–216. [http://doi.org/10.1016/s0165-0327\(00\)00338-4](http://doi.org/10.1016/s0165-0327(00)00338-4)
- Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 33(2), 81–88. <http://doi.org/10.1016/j.neubiorev.2008.08.004>
- Åhs, F., Sollers, J. J., Furmark, T., Fredrikson, M., & Thayer, J. F. (2009). High-frequency heart rate variability and cortico-striatal activity in men and women with social phobia. *NeuroImage*, 47(3), 815–820. <http://doi.org/10.1016/j.neuroimage.2009.05.091>
- Napadow, V., Dhond, R., Conti, G., Makris, N., Brown, E. N., & Barbieri, R. (2008). Brain correlates of autonomic modulation: Combining heart rate variability with fMRI. *NeuroImage*, 42(1), 169–177. <http://doi.org/10.1016/j.neuroimage.2008.04.238>
- Smith, R., Allen, J. J. B., Thayer, J. F., Fort, C., & Lane, R. D. (2014). Increased association over time between regional frontal lobe BOLD change magnitude and cardiac vagal control with sertraline treatment for major depression. *Psychiatry Research: Neuroimaging*, 224(3), 225–233. <http://doi.org/10.1016/j.psychresns.2014.08.015>
- Winkelmann, T., Thayer, J. F., Pohlack, S., Nees, F., Grimm, O., & Flor, H. (2016). Structural brain correlates of heart rate variability in a healthy young adult population. *Brain Structure and Function*, 222(2), 1061–1068. <http://doi.org/10.1007/s00429-016-1185-1>

- Adhikari, A., Lerner, T. N., Finkelstein, J., Pak, S., Jennings, J. H., Davidson, T. J., ... Deisseroth, K. (2015). Basomedial amygdala mediates top-down control of anxiety and fear. *Nature*, *527*(7577), 179–185. <http://doi.org/10.1038/nature15698>
- Sakaki, M., Yoo, H. J., Nga, L., Lee, T.-H., Thayer, J. F., & Mather, M. (2016). Heart rate variability is associated with amygdala functional connectivity with MPFC across younger and older adults. *NeuroImage*, *139*, 44–52. <http://doi.org/10.1016/j.neuroimage.2016.05.076>
- Lane, R., McRae, K., Reiman, E., Chen, K., Ahern, G., & Thayer, J. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage*, *44*(1), 213–222. <http://doi.org/10.1016/j.neuroimage.2008.07.056>
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A Polyvagal Theory. *Psychophysiology*, *32*(4), 301–318. <http://doi.org/10.1111/j.1469-8986.1995.tb01213.x>
- Porges, S. W. (2001). The polyvagal theory: phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, *42*(2), 123–146. [http://doi.org/10.1016/s0167-8760\(01\)00162-3](http://doi.org/10.1016/s0167-8760(01)00162-3)
- Porges, S. (2011). *The polyvagal theory : neurophysiological foundations of emotions, attachment, communication, and self-regulation*. New York: W.W. Norton.
- Levy, B. J., & Anderson, M. C. (2012). Purging of Memories from Conscious Awareness Tracked in the Human Brain. *Journal of Neuroscience*, *32*(47), 16785–16794. <http://doi.org/10.1523/jneurosci.2640-12.2012>
- van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory*, *25*(9), 1201–1216. <http://doi.org/10.1080/09658211.2017.1282518>
- Hellerstedt, R., Johansson, M., & Anderson, M. C. (2016). Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia*, *89*, 510–523. <http://doi.org/10.1016/j.neuropsychologia.2016.07.008>
- Hu, X., Bergström, Z. M., Gagnepain, P., & Anderson, M. C. (2017). Suppressing Unwanted Memories Reduces Their Unintended Influences. *Current Directions in Psychological Science*, *26*(2), 197–206. <http://doi.org/10.1177/0963721417689881>

- Lambert, A. J., Good, K. S., & Kirk, I. J. (2010). Testing the repression hypothesis: Effects of emotional valence on memory suppression in the think – No think task. *Consciousness and Cognition*, *19*(1), 281–293. <http://doi.org/10.1016/j.concog.2009.09.004>
- Medina, A. M., Clark, M., & Thorne, G. (2016). Disgust and Memory for Visually and Verbally Presented Information.. *North American Journal of Psychology*, *18*(3), 571–576. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=bsx&AN=119642011&lang=fr&authtype=ip,uid>
- Newhagen, J. E. (1998). TV news images that induce anger fear, and disgust: Effects on approach-avoidance and memory. *Journal of Broadcasting & Electronic Media*, *42*(2), 265–276. <http://doi.org/10.1080/08838159809364448>
- Rohrman, S., & Hopp, H. (2008). Cardiovascular indicators of disgust. *International Journal of Psychophysiology*, *68*(3), 201–208. <http://doi.org/10.1016/j.ijpsycho.2008.01.011>
- Mccraty, R., & Shaffer, F. (2015). Heart Rate Variability: New Perspectives on Physiological Mechanisms Assessment of Self-regulatory Capacity, and Health Risk. *Global Advances in Health and Medicine*, *4*(1), 46–61. <http://doi.org/10.7453/gahmj.2014.073>
- Pascalis, V. D., Barry, R. J., & Sparita, A. (1995). Decelerative changes in heart rate during recognition of visual stimuli: effects of psychological stress. *International Journal of Psychophysiology*, *20*(1), 21–31. [http://doi.org/10.1016/0167-8760\(95\)00023-1](http://doi.org/10.1016/0167-8760(95)00023-1)
- Abercrombie, H. C., Chambers, A. S., Greischar, L., & Monticelli, R. M. (2008). Orienting emotion, and memory: Phasic and tonic variation in heart rate predicts memory for emotional pictures in men. *Neurobiology of Learning and Memory*, *90*(4), 644–650. <http://doi.org/10.1016/j.nlm.2008.08.001>
- Ruiz-Padial, E., Medialdea, M. T. M., del Paso, G. R., & Thayer, J. F. (2017). Individual Differences in Attentional Capture by Pictures of Fear and Disgust as Indexed by Cardiac Responses. *Journal of Psychophysiology*, 1–11. <http://doi.org/10.1027/0269-8803/a000198>
- Benoit, R. G., Davies, D. J., & Anderson, M. C. (2016). Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences*, *113*(52), E8492–E8501. <http://doi.org/10.1073/pnas.1606604114>
- Ketz, N., O'Reilly, R. C., & Curran, T. (2014). Classification aided analysis of oscillatory

signatures in controlled retrieval.. *Neuroimage*, 85 Pt 2, 749–60.

Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and Neural Oscillations during Volitional Suppression of Memory Retrieval. *Journal of Cognitive Neuroscience*, 25(10), 1624–1633. http://doi.org/10.1162/jocn_a_00418

Waldhauser, G. T., Bäuml, K.-H. T., & Hanslmayr, S. (2014). Brain Oscillations Mediate Successful Suppression of Unwanted Memories. *Cerebral Cortex*, 25(11), 4180–4190. <http://doi.org/10.1093/cercor/bhu138>

Lambert, A. J., Good, K. S., & Kirk, I. J. (2010). Testing the repression hypothesis: Effects of emotional valence on memory suppression in the think – No think task. *Consciousness and Cognition*, 19(1), 281–293. <http://doi.org/10.1016/j.concog.2009.09.004>

Vito, D. D., & Fenske, M. J. (2017). Suppressing memories of words and familiar objects results in their affective devaluation: Evidence from Think/No-think tasks. *Cognition*, 162, 1–11. <http://doi.org/10.1016/j.cognition.2017.01.020>

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, 48(4), 726–737. <http://doi.org/10.1016/j.neuroimage.2009.06.051>

Bergström, Z. M., de Fockert, J., & Richardson-Klavehn, A. (2009). Event-related Potential Evidence that Automatic Recollection Can Be Voluntarily Avoided. *Journal of Cognitive Neuroscience*, 21(7), 1280–1301. <http://doi.org/10.1162/jocn.2009.21075>

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, 94(1), 1–11. <http://doi.org/10.1016/j.biopsycho.2013.04.012>

Bergström, Z. M., Velmans, M., de Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. *Brain Research*, 1151, 119–133. <http://doi.org/10.1016/j.brainres.2007.03.014>

Chen, C., Liu, C., Huang, R., Cheng, D., Wu, H., Xu, P., ... Luo, Y.-J. (2012). Suppression of aversive memories associates with changes in early and late stages of neurocognitive processing. *Neuropsychologia*, 50(12), 2839–2848. <http://doi.org/10.1016/j.neuropsychologia.2012.08.004>

- Hanslmayr, S., Leopold, P., Pastotter, B., & Bauml, K.-H. (2009). Anticipatory Signatures of Voluntary Memory Suppression. *Journal of Neuroscience*, *29*(9), 2742–2747. <http://doi.org/10.1523/jneurosci.4703-08.2009>
- Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain Research*, *1255*, 132–147. <http://doi.org/10.1016/j.brainres.2008.11.073>
- Heusser, A. C., Poeppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta–gamma phase code. *Nature Neuroscience*, *19*(10), 1374–1380. <http://doi.org/10.1038/nn.4374>
- Clouter, A., Shapiro, K. L., & Hanslmayr, S. (2017). Theta Phase Synchronization Is the Glue that Binds Human Associative Memory. *Current Biology*, *27*(20), 3143–3148.e6. <http://doi.org/10.1016/j.cub.2017.09.001>
- Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neurosci Biobehav Rev*, *34*, 1023–35.
- Burke, J. F., Sharan, A. D., Sperling, M. R., Ramayya, A. G., Evans, J. J., Healey, M. K., ... Kahana, M. J. (2014). Theta and High-Frequency Activity Mark Spontaneous Recall of Episodic Memories. *Journal of Neuroscience*, *34*(34), 11355–11365. <http://doi.org/10.1523/jneurosci.2654-13.2014>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. <http://doi.org/10.1016/j.tics.2014.04.012>
- Mückschel, M., Dippel, G., & Beste, C. (2017). Distinguishing stimulus and response codes in theta oscillations in prefrontal areas during inhibitory control of automated responses. *Human Brain Mapping*, *38*(11), 5681–5690. <http://doi.org/10.1002/hbm.23757>
- Eschmann, K. C. J., Bader, R., & Mecklinger, A. (2018). Topographical differences of frontal-midline theta activity reflect functional differences in cognitive control abilities. *Brain and Cognition*, *123*, 57–64. <http://doi.org/10.1016/j.bandc.2018.02.002>
- Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and neural oscillations during volitional suppression of memory retrieval. *J Cogn Neurosci*, *25*, 1624–33.

- Ketz, N., O'Reilly, R. C., & Curran, T. (2014). Classification aided analysis of oscillatory signatures in controlled retrieval. *NeuroImage*, *85*, 749–760. <http://doi.org/10.1016/j.neuroimage.2013.06.077>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. <http://doi.org/10.1016/j.jneumeth.2007.03.024>
- Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology*, *3*. <http://doi.org/10.3389/fpsyg.2012.00606>
- Ottaviani, C., Zingaretti, P., Petta, A. M., Antonucci, G., Thayer, J. F., & Spitoni, G. F. (2018). Resting Heart Rate Variability Predicts Inhibitory Control Above and Beyond Impulsivity. *Journal of Psychophysiology*, 1–9. <http://doi.org/10.1027/0269-8803/a000222>
- Gillie, B. L., Vasey, M. W., & Thayer, J. F. (2015). Individual differences in resting heart rate variability moderate thought suppression success. *Psychophysiology*, *52*(9), 1149–1160. <http://doi.org/10.1111/psyp.12443>
- Gillie, B. L., Vasey, M. W., & Thayer, J. F. (2013). Heart Rate Variability Predicts Control Over Memory Retrieval. *Psychological Science*, *25*(2), 458–465. <http://doi.org/10.1177/0956797613508789>
- Riegel, M., Łukasz Żurawski, Wierzba, M., Moslehi, A., Łukasz Klocek, Horvat, M., ... Marchewka, A. (2015). Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). *Behavior Research Methods*, *48*(2), 600–612. <http://doi.org/10.3758/s13428-015-0620-1>
- Davies, M. I., & Clark, D. M. (1998). Thought suppression produces a rebound effect with analogue post-traumatic intrusions. *Behaviour Research and Therapy*, *36*(6), 571–582. [http://doi.org/10.1016/s0005-7967\(98\)00051-5](http://doi.org/10.1016/s0005-7967(98)00051-5)
- Shipherd, J. C., & Beck, J. G. (2005). The role of thought suppression in posttraumatic stress disorder. *Behavior Therapy*, *36*(3), 277–287. [http://doi.org/10.1016/s0005-7894\(05\)80076-0](http://doi.org/10.1016/s0005-7894(05)80076-0)
- Klimesch, W. (2012). Alpha-band oscillations attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. <http://doi.org/10.1016/j.tics.2012.10.007>

- Sakaki, M., Yoo, H. J., Nga, L., Lee, T.-H., Thayer, J. F., & Mather, M. (2016). Heart rate variability is associated with amygdala functional connectivity with MPFC across younger and older adults. *NeuroImage*, *139*, 44–52. <http://doi.org/10.1016/j.neuroimage.2016.05.076>
- Gold, A. L., Morey, R. A., & McCarthy, G. (2015). Amygdala–Prefrontal Cortex Functional Connectivity During Threat-Induced Anxiety and Goal Distraction. *Biological Psychiatry*, *77*(4), 394–403. <http://doi.org/10.1016/j.biopsych.2014.03.030>
- Wei, L., Chen, H., & Wu, G.-R. (2018). Structural Covariance of the Prefrontal-Amygdala Pathways Associated with Heart Rate Variability. *Frontiers in Human Neuroscience*, *12*. <http://doi.org/10.3389/fnhum.2018.00002>
- Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, *36*(2), 747–756. <http://doi.org/10.1016/j.neubiorev.2011.11.009>
- Adhikari, A., Lerner, T. N., Finkelstein, J., Pak, S., Jennings, J. H., Davidson, T. J., ... Deisseroth, K. (2015). Basomedial amygdala mediates top-down control of anxiety and fear. *Nature*, *527*(7577), 179–185. <http://doi.org/10.1038/nature15698>
- Yang, T. T., Simmons, A. N., Matthews, S. C., Tapert, S. F., Bischoff-Grethe, A., Frank, G. K. W., ... Paulus, M. P. (2007). Increased amygdala activation is related to heart rate during emotion processing in adolescent subjects. *Neuroscience Letters*, *428*(2-3), 109–114. <http://doi.org/10.1016/j.neulet.2007.09.039>
- Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160206. <http://doi.org/10.1098/rstb.2016.0206>
- Rohrman, S., Hopp, H., Schienle, A., & Hodapp, V. (2009). Emotion regulation disgust sensitivity, and psychophysiological responses to a disgust-inducing film. *Anxiety, Stress & Coping*, *22*(2), 215–236. <http://doi.org/10.1080/10615800802016591>
- Broderick, J., Grisham, J. R., & Weidemann, G. (2013). Disgust and Fear Responding in Contamination-Based Obsessive-Compulsive Disorder During Pictorial Exposure. *Behavior Therapy*, *44*(1), 27–38. <http://doi.org/10.1016/j.beth.2012.05.005>
- Shenhav, A., & Mendes, W. B. (2014). Aiming for the stomach and hitting the heart:

- Dissociable triggers and sources for disgust reactions.. *Emotion*, *14*(2), 301–309. <http://doi.org/10.1037/a0034644>
- Stark, R., Walter, B., Schienle, A., & Vaitl, D. (2005). Psychophysiological Correlates of Disgust and Disgust Sensitivity. *Journal of Psychophysiology*, *19*(1), 50–60. <http://doi.org/10.1027/0269-8803.19.1.50>
- Gilchrist, P. T., Vranceanu, T., Béland, S., Bacon, S. L., & Ditto, B. (2016). Disgust stimuli reduce heart rate but do not contribute to vasovagal symptoms. *Journal of Behavior Therapy and Experimental Psychiatry*, *51*, 116–122. <http://doi.org/10.1016/j.jbtep.2016.01.005>
- Meissner, K., Muth, E. R., & Herbert, B. M. (2011). Bradygastric activity of the stomach predicts disgust sensitivity and perceived disgust intensity. *Biological Psychology*, *86*(1), 9–16. <http://doi.org/10.1016/j.biopsycho.2010.09.014>
- Olatunji, B. O., Armstrong, T., & Elwood, L. (2017). Is Disgust Proneness Associated With Anxiety and Related Disorders? A Qualitative Review and Meta-Analysis of Group Comparison and Correlational Studies. *Perspectives on Psychological Science*, *12*(4), 613–648. <http://doi.org/10.1177/1745691616688879>
- Cisler, J. M., Olatunji, B. O., & Lohr, J. M. (2009). Disgust fear, and the anxiety disorders: A critical review. *Clinical Psychology Review*, *29*(1), 34–46. <http://doi.org/10.1016/j.cpr.2008.09.007>
- Brady, R. E., Adams, T. G., & Lohr, J. M. (2010). Disgust in contamination-based obsessive-compulsive disorder: a review and model. *Expert Review of Neurotherapeutics*, *10*(8), 1295–1305. <http://doi.org/10.1586/ern.10.46>
- Vicario, C. M. (2013). Altered Insula Response to Sweet Taste Processing in Recovered Anorexia and Bulimia Nervosa: a Matter of Disgust Sensitivity?. *American Journal of Psychiatry*, *170*(12), 1497–1497. <http://doi.org/10.1176/appi.ajp.2013.13060748>
- Vicario, C. M., Rafal, R. D., Martino, D., & Avenanti, A. (2017). Core social and moral disgust are bounded: A review on behavioral and neural bases of repugnance in clinical disorders. *Neuroscience & Biobehavioral Reviews*, *80*, 185–200. <http://doi.org/10.1016/j.neubiorev.2017.05.008>
- Powell, P. A., Overton, P. G., & Simpson, J. (2013). The Revolting Self: An Interpretative Phenomenological Analysis of the Experience of Self-Disgust in Females With Depressive

Symptoms. *Journal of Clinical Psychology*, 70(6), 562–578. <http://doi.org/10.1002/jclp.22049>

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS) a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, 5(5), e10773. <http://doi.org/10.1371/journal.pone.0010773>

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, 9(9), e106953. <http://doi.org/10.1371/journal.pone.0106953>

Marchewka, A., Łukasz Żurawski, Jednoróg, K., & Grabowska, A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, 46(2), 596–610. <http://doi.org/10.3758/s13428-013-0379-1>

Wierzbna, M., Riegel, M., Pucz, A., Leśniewska, Z., Dragan, W. L., Gola, M., . . . Marchewka, A. (2015). Erotic subset for the Nencki Affective Picture System (NAPS ERO): cross-sexual comparison study. *Frontiers in Psychology*, 6. <http://doi.org/10.3389/fpsyg.2015.01336>

Wierzbna, M., Riegel, M., Wypych, M., Jednoróg, K., Turnau, P., Grabowska, A., & Marchewka, A. (2015). Basic Emotions in the Nencki Affective Word List (NAWL BE): New Method of Classifying Emotional Stimuli. *PLOS ONE*, 10(7), e0132305. <http://doi.org/10.1371/journal.pone.0132305>

Lang, P. J. (2008). International affective picture system (IAPS): affective ratings of pictures and instruction manual. *Technical Report A-8*. Gainesville, FL: University of Florida.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net.. *Electroencephalogr Clin Neurophysiol*, 87, 154–63.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality.. *Clin Neurophysiol*, 112, 536–44.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7. <http://doi.org/10.3389/fnins.2013.00267>

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . .

- Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. <http://doi.org/10.1016/j.neuroimage.2013.10.027>
- Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. <http://doi.org/10.3389/fninf.2015.00016>
- Andersen, L. M. (2018). Group Analysis in MNE-Python of Evoked Responses from a Tactile Stimulation Paradigm: A Pipeline for Reproducibility at Every Step of Processing Going from Individual Sensor Space Representations to an across-Group Source Space Representation. *Frontiers in Neuroscience*, *12*. <http://doi.org/10.3389/fnins.2018.00006>
- Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, *12*. <http://doi.org/10.3389/fnins.2018.00530>
- Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. <http://doi.org/10.1016/j.neuroimage.2017.06.030>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *afex: Analysis of Factorial Experiments*. Retrieved from <https://CRAN.R-project.org/package=afex>
- Torchiano, M. (2017). *effsize: Efficient Effect Size Computation*. Retrieved from <https://CRAN.R-project.org/package=effsize>

3.1 Les dynamiques oscillatoires de la suppression mnésique

Le contrôle volontaire de souvenirs négatifs reste décrit - tant par les individus qui participent à ces épreuves que par un certain lieu commun - comme une tâche difficile et coûteuse cognitivement, si ce n'est impossible. A ceci s'ajoute même la notion, encore très prégnante sur la psychologie et la psychiatrie d'obédience psychanalytique, selon laquelle la suppression directe de pensées ou de souvenirs se traduit systématiquement par un effet de rebond, de retour du refoulé, ce qui la disqualifierait de toute ambition clinique sérieuse. C'est pourtant dans ce domaine que la compréhension des dynamiques du contrôle cognitif est la plus nécessaire, les intrusions apparaissant en lien avec de nombreux symptômes comme une plus grande tendance à la rumination (Disner et al., 2011), les troubles anxieux, les états dépressifs (Mihailova and Jobson, 2018; Brewin et al., 2010), le trouble de stress post-traumatique (Brewin, 2014) ou encore les troubles obsessionnels compulsifs (Speckens et al., 2007). Nous avons montré dans l'étude précédente que, lorsqu'il est effectif, le contrôle inhibiteur des souvenirs est associé à ce qui peut être interprété comme un signe de régulation des composantes émotionnelles du souvenir, et que ce n'est qu'en cas de défaillance de ce contrôle qu'un effet de type "rebond" peut s'observer. Il convient toutefois d'examiner plus avant ce qu'est réellement l'inhibition d'une intrusion, dans cette étude ou d'autres ayant abordé le paradigme "Think/No-Think", et les composantes qui permettent un oubli. Dans la partie suivante, nous allons tenter d'examiner en détail la séquence sous-tendant l'intrusion de souvenirs et les dynamiques oscillatoires qui sous-tendent la suppression directe de celle-ci. Il nous faudra au préalable dépasser trois notions qui entourent l'idée d'oubli volontaire : **1)** qu'il s'agit d'une mécanique énigmatique dans sa réalisation, qu'on ne pourrait ni apprendre ni entraîner, **2)** que le contrôle des pensées est associé à des contre manifestations négatives et que, lorsqu'il réussit, **3)** il s'agit en réalité d'un phénomène lié à la distraction ou à la réévaluation de l'importance des souvenirs qu'il a induit et que la distraction ou l'évitement sont les véritables clés de l'amélioration observée du contrôle des intrusions.

3.1.1 Intrusions mnésiques et psychopathologies

L'oubli est plus que l'affaiblissement passif des traces de la mémoire. Il s'agit d'un processus actif, et réactif. Il peut intervenir en contrecoup d'éléments de l'environnement, mais son succès potentiel peut aussi dépendre de mécanismes qui sont initiés volontairement par l'individu (Levy and Anderson, 2008). Ceci est particulièrement vrai pour ce qui concerne les formes d'oublis que nous avons décrites comme des réactions à un contexte spécifique, qu'il s'agisse de l'oubli induit

par la récupération de souvenirs semblables (Anderson et al., 1994), de l'oubli dirigé qui interfère avec l'encodage de la trace mnésique (MacLeod, 1975) ou encore de l'oubli volontaire qui nous intéresse ici et agit par inhibition de la récupération du souvenir (Anderson and Green, 2001). Dans ce dernier cas, on considère que l'inhibition du rappel est un mécanisme initié en réaction à une réactivation non intentionnelle d'événements dont le rappel interfère avec d'autres objectifs poursuivis par l'individu (Anderson, 2003). L'inhibition du rappel involontaire d'événements traumatisants dans le cadre du syndrome de stress post-traumatique est une illustration d'un tel besoin. Alors même qu'il est aujourd'hui compris que la régulation inhibitrice qui est engagée en réaction aux intrusions lors de la tâche de "Think/No-Think" induit à la fois une réduction de celles-ci et un oubli progressif des souvenirs ciblés (Benoit et al., 2015; Schie and Anderson, 2017), on observe généralement une grande variabilité entre les individus quant à la réussite de ces mécanismes. Il semble même que cette capacité soit influencée par des composantes environnementales et qu'elle puisse varier au cours du temps chez une même personne. Dans une étude récente, Hulbert and Anderson, 2018 ont par exemple montré que des participants ayant été exposés à un niveau modéré d'expériences traumatisantes dans leur vie montrent de meilleures capacités de suppression mnésique lors d'épreuves de "Think/No-Think" que des participants qui rapportaient peu ou pas de présence de traumas. D'une certaine manière, l'exposition à un niveau non psychopathogène d'expérience négative semble avoir induit le développement plus important des fonctions de résilience qui pourraient donc, à l'image d'autres fonctions cognitives, faire l'objet d'un entraînement et s'améliorer par la pratique. (Lyoo, 2011) a par exemple rapporté qu'un an après la survenue d'un événement traumatisant, l'épaisseur corticale des zones préfrontales latérales - qui recourent les substrats de la suppression volontaires des souvenirs - avait augmenté chez les personnes exposées à ces événements. Cet effet de l'exposition, qui pourrait sembler bénéfique au regard des capacités d'inhibition, n'est pourtant pas linéaire : les personnes présentant une dépression majeure (Joormann et al., 2009), des traumas trop importants (Waldhauser et al., 2018) ou un trouble de stress post-traumatique persistant (Catarino et al., 2015) ont malgré tout de plus grandes difficultés que des sujets contrôles. Mais ces résultats montrent que les capacités d'inhibition sur les souvenirs apparaissent en réaction à l'augmentation des intrusions. L'hypothèse selon laquelle un certain nombre de psychopathologies pourraient être liées non pas à de trop faibles capacités de contrôle, mais à une insuffisante adaptation de celle-ci en réaction au trauma est donc une alternative considérée sérieusement.

Ces évolutions des capacités d'inhibition et de leurs substrats cérébraux en fonction des traumas auxquels sont confrontés les individus, et le lien que ces dynamiques entretiennent avec le développement des psychopathologies, ne changent pas toutefois la difficulté et le coût cognitif qui semblent associés à l'oubli volontaire. L'oubli, quand il n'est pas passif, est perçu comme une épreuve. Umberto Eco considérait qu'il était fondamentalement impossible (Eco and Migiel, 1988). La première difficulté que rencontre toute tentative d'oubli volontaire se trouve précisément dans le lien

intime qu'entretient l'acte même de la volonté d'oublier avec son objet. On pourrait difficilement parler d'une intentionnalité de l'oubli, prise au sens husserlien comme une "action de tendre vers" (*intentio*) propre à cette volonté de suppression : si la conscience est toujours conscience de quelque chose, l'oubli comme processus cognitif doit déjà se défaire de ce lien et ne pas être l'oubli "de..." sous peine de reprendre en permanence ce qu'il cherche à occulter. C'est ce que notait par exemple Jean de La Bruyère dans *Les Caractères* (Bruyère, 1688/1995) dans la section consacrée justement au Cœur : "Il faut, s'il se peut, ne point songer à sa passion pour l'affaiblir", disait-il en un sens qui n'est pas différent de ce que nous avons pu évoquer dans la partie précédente mais aussi, un peu plus haut, "Vouloir oublier quelqu'un, c'est y penser".

Ce paradoxe apparent a été alimenté sur le plan de la psychologie cognitive par de nombreuses expériences concernant la suppression des pensées. Les observations qui ont entouré la célèbre procédure proposée par (Wegner, 1994) ("*Ne pensez pas à l'ours blanc*") ont par exemple été souvent interprétées comme des preuves de la nocivité, sinon de l'impossibilité des tentatives de suppression des pensées. Dans cette procédure, il est demandé aux participants de supprimer de leurs pensées la représentation visuelle d'un ours blanc. Les résultats montrent notamment un effet paradoxal de la tentative de contrôle dans la mesure où la représentation a tendance à revenir davantage à l'esprit avec notamment un effet de rebond (Wegner et al., 1987). Mais ces données ne permettent pas tout à fait de conclure sur l'utilité de la suppression directe dans le cas d'intrusions mentales. Elles montrent davantage une difficulté certaine à maintenir en mémoire de travail les éléments d'une instruction pour un test sans que ceux-ci ne réapparaissent entièrement à la conscience. Il faut ainsi distinguer ce mécanisme de "suppression des pensées", dont l'existence n'est pas discutée ici, de la suppression volontaire telle qu'elle est abordée via le paradigme "*Think/No-Think*" dans la mesure où la représentation n'est, dans ce dernier cas, pas activement maintenue à l'esprit par les instructions même de la tâche. En particulier, la difficulté rencontrée dans la procédure de (Wegner, 1994) ne repose pas sur des déficits des capacités d'inhibition mnésique car il n'existe pas de différences probantes sur le plan clinique entre des patients et des sujets contrôles en terme de performances à ce test (Magee et al., 2012). Un autre argument assez proche qui peut être avancé souligne que la stratégie même de suppression des pensées est en général associée à une plus forte prévalence de psychopathologies, notamment à travers la comparaison de profils de réponses à des questionnaires évaluant la tendance à vouloir supprimer les pensées. Il est pourtant apparu que ces questionnaires (e.g. *White Bear Suppression Inventory*) évaluent autant la tendance à vouloir supprimer les pensées que les échecs perçus de ces tentatives et la réponse émotionnelle qui en résulte (Höping and Jong-Meyer, 2003). La propension d'un individu à supprimer volontairement des pensées lui paraîtra ainsi d'autant plus important que ses échecs sont nombreux et associés à un ressenti émotionnel ou des jugements négatifs sur ses propres compétences.

Les patients ont souvent parfaitement conscience de leurs difficultés et de l'influence que ces intrusions peuvent avoir dans l'adaptation de leurs comportements. Pour les contenir, des stratégies d'évitement ou de distraction sont souvent utilisées. Pourtant, celles-ci ont des effets potentiellement paradoxaux comme l'augmentation du nombre futur des intrusions, une hyper-vigilance aux indices dans l'environnement qui les déclenchent et une évaluation affective encore plus négative lors de leur apparition (Purdon, 2004). A nouveau, ces stratégies d'évitement ou de distraction doivent être dissociées de la suppression volontaire que nous étudions ici. L'oubli induit par la substitution de pensée (*thought substitution*), peut être observé lorsqu'il est demandé au participant de récupérer volontairement un souvenir alternatif lorsqu'un indice évoquant le souvenir cible est présenté. Rapporté au contexte du paradigme "Think/No-Think", cela revient à demander au participant de rappeler consciemment un souvenir lorsqu'il est confronté à un indice de type "No-Think". Cette procédure induit effectivement une forme d'oubli par le biais d'un mécanisme qui est assez proche de l'oubli induit par la récupération : le souvenir cible est inhibé pour permettre la ré-évoation d'un souvenir alternatif (Wimber et al., 2015; Waldhauser et al., 2012). Mais ce processus est différent de l'inhibition directe sans distraction, il fait notamment intervenir des aires cérébrales différentes et induit des effets quantitativement moins importants (Benoit and Anderson, 2012). Pourtant, l'hypothèse même de la suppression directe n'a que rarement été évaluée en tant que telle par les recherches précédentes qui dissociaient peu les stratégies d'évitement de l'inhibition directe des intrusions (Catarino et al., 2015).

Les modèles neurobiologiques actuels de la suppression directe des souvenirs (Hu et al., 2017; Anderson et al., 2011; Anderson et al., 2016) se réfèrent généralement à l'idée selon laquelle le contrôle des intrusions correspond à un processus réactionnel à la détection corticale d'une activation hippocampique (Anderson, 2003), induisant un signal inhibiteur en provenance du cortex préfrontal latéral droit (Anderson, 2004; Depue et al., 2007; Gagnepain et al., 2014; Gagnepain et al., 2017). Ce signal est la composante essentielle de l'oubli volontaire. On considère qu'il peut influencer l'activité du lobe temporal médian bien au-delà de l'apparition de l'intrusion, et induire une forme d'amnésie des événements adjacents dans le temps (Hulbert et al., 2016). Ainsi, chercher à supprimer volontairement un souvenir peut être suffisamment puissant pour altérer jusqu'à l'encodage de nouveaux souvenirs issus des expériences récentes. La mise en évidence de ce phénomène a évidemment une grande importance pour la compréhension de l'altération de la mémoire pour des personnes souffrant de traumatismes chez qui les tentatives de contrôles sont récurrentes.

Les dynamiques cognitives qui entourent l'apparition de ces intrusions sont donc centrales si l'on veut saisir les défenses opposées aux intrusions et leur effet sur le long terme. La nature du signal qui est envoyé à l'hippocampe reste un point crucial sur lequel nous n'avons que peu de visibilité. De telles réactions sont par définition très difficiles à déceler lors de paradigmes en imagerie fonctionnelle dont la résolution temporelle ne permet que l'étude de phénomènes relativement lents

(> 1-2 seconde). L'analyse d'enregistrements en électroencéphalographie (EEG), en magnétoencéphalographie (MEG) ou l'enregistrement direct de populations de neurones par électro-encéphalographie intracrânienne (iEEG) - placée à la surface du cortex ou dans certaines zones sous-corticales - offrent à l'inverse une précision plus à même de révéler ces patterns d'activité. Dans cette section, nous allons essayer de comprendre la séquence de processus qui permet une réaction efficace des capacités de contrôle face à l'intrusion non désirée d'un souvenir. Mais puisque ces intrusions semblent imprévisibles, tant pour le sujet que pour tout observateur extérieur, une première tâche sera d'en comprendre les mécanismes d'apparition et comment les mesurer.

3.1.2 Séquence oscillatoire de l'inhibition mnésique

Les oscillations neurales qui sont enregistrées en électrophysiologie reflètent les fluctuations des potentiels de champs locaux générés par les phases excitatrices et inhibitrices des réseaux de neurones (Cohen, 2017). Les changements observables de ces oscillations sont des marqueurs importants de l'implémentation neurale des processus cognitifs (Schmidt et al., 2018). Le tracé EEG qui est obtenu lorsque l'activité électrique est enregistrée à la surface du crâne peut notamment être caractérisé en terme d'oscillations de différentes fréquences. Ces oscillations correspondent à des patterns d'activité répétitifs assimilables à des courbes sinusoïdales reproduisant leur cycle plusieurs fois par seconde. Des analyses consacrées (transformée de Fourier ou de Hilbert, ondelettes de Morlet...) permettent de quantifier l'intensité de différentes bandes de fréquence et éventuellement leur évolution dans le temps. On distingue alors une activité Delta inférieure à 4 Hz de l'activité Thêta (3-8 Hz), Alpha (8-12 Hz) ou Bêta (13-30 Hz). Nous ne tenterons pas ici d'associer de façon trop réductrice les puissances de ces bandes de fréquence, qui sont les composants élémentaires d'une dynamique physiologique complexe, avec des états mentaux élaborés qui pourraient souvent être divisés en sous processus cognitifs (e.g. la veille, le sommeil, l'imagination, etc...). Il importe malgré tout de comprendre ce que peuvent signifier les augmentations ou diminutions brusques de l'amplitude de ces fréquences dans le contexte de tâches bien déterminées.

Nous nous intéresserons en particulier ici à la bande de fréquence Thêta (entre environs 3 et 8 Hz chez l'Homme et 4 et 12 Hz lorsqu'elle est enregistrée *in vivo* dans l'hippocampe de rats). Cette fréquence est l'une des plus basse qui soit analysée sur le tracé EEG et offre l'avantage de présenter des oscillations de larges amplitudes, ce qui en rend généralement l'observation plus aisée et plus résistante au bruit. Mais ces oscillations ont aussi été identifiées pour leur implication dans la synchronisation d'aires cérébrales proches ou distantes (Clouter et al., 2017). Un nombre important d'études a notamment rapporté son implication dans des processus de contrôle inhibiteur tels qu'ils nous intéressent ici (Chmielewski et al., 2015; Babapoor-Farrokhran et al., 2017; Cavanagh and Frank, 2014), mais aussi pour l'encodage et le rappel de souvenirs (Anderson et al., 2009; Hanslmayr et al., 2012; Burke et al., 2014; Fuentemilla et al., 2014), bien qu'il existe de nombreuses divergences quant à l'augmentation ou à la réduction de cette bande de fréquence

dans ce contexte. Plus spécifiquement, l'activité Thêta enregistrée au niveau des électrodes frontales médianes, qui sont fortement influencées par l'activité des neurones présents dans le cortex préfrontal médian, semble refléter l'implémentation du contrôle de comportements orientés vers un but lorsque des réactions automatisées sont mobilisées (Cavanagh and Frank, 2014). Ces oscillations joueraient ainsi un rôle de métronome pour favoriser la communication au sein même du cortex frontal et avec des aires distantes à l'origine des conflits cognitifs détectés. Des études en potentiels évoqués ont montré un lien avec le besoin de contrôle (Walsh and Anderson, 2012; Folstein and Petten, 2007), et il en est de même pour les approches fréquentielles (Smit et al., 2005). Ces augmentations apparaissent notamment lorsque des processus entrent en conflit (Hanslmayr et al., 2008; Nigbur et al., 2011; Frank et al., 2015) et l'augmentation de cette bande de fréquence semble implémenter la résolution de ceux-ci via un contrôle cognitif (Verguts, 2017).

3.1.3 Analyses de patterns multivariées en électroencéphalographie pour la détection d'intrusions

Compte-tenu des arguments fournis par la littérature existante, nous avons choisi d'orienter nos hypothèses vers la mesure de cette bande de fréquence comme marqueur de la détection d'intrusions mnésiques. L'analyse des enregistrements réalisés en électrophysiologie consiste souvent, dans le cadre d'approches neurocognitives, à contraster l'activité induite par deux conditions expérimentales. Les potentiels évoqués sont une approche possible qui utilise le moyennage d'un grand nombre d'enregistrements liés aux différentes conditions pour comparer l'évolution de l'activité électrique au cours du temps. Ce type d'approche, qui trouve également son pendant dans le cadre d'études en IRM fonctionnelle, utilise les paramètres statistiques des distributions observées (e.g. une différence de moyennes) pour inférer l'existence de niveaux d'activation différents et peut être qualifié d'**univarié**. Le **décodage multivarié** des états mentaux a été proposé comme une approche alternative et se trouve de plus en plus utilisé dans le cadre d'études en neurosciences cognitives (Grootswagers et al., 2017; Hebart and Baker, 2018; Wallis, 2018). Ces approches intègrent, au sein d'un même modèle, l'ensemble des mesures réalisées (l'ensemble des électrodes ou l'ensemble des voxels s'il s'agit d'un paradigme en imagerie fonctionnelle), ce qui permet d'augmenter à la fois la sensibilité et la spécificité des fonctions qui sont inférées en tenant compte des variations mais aussi des covariations qui permettent de les déceler. Cette différence a notamment l'avantage de rendre l'apprentissage automatisé plus sensible aux effets explorés (Norman et al., 2006) mais aussi plus spécifique quant aux variables qu'il met en lumière (Haynes, 2015), ce qui peut en partie expliquer son utilisation pour l'approche d'états mentaux complexes ou difficilement discernables au sein des données recueillies.

Les techniques de classifications multivariées utilisent des algorithmes développés pour la reconnaissance de patterns au sein de données à grandes dimensions comme des images et qui sont longtemps restées spécifiques aux études liées aux interfaces

cerveau-machines. Les performances prédictives de tels modèles (e.g. reconnaître des chiffres manuscrits ou les expressions d'émotions sur le visage d'un individu) étaient, dans ce cadre là, la métrique la plus importante (Müller et al., 2008). Mais l'application, plus récente, de ces méthodes au champ des neurosciences cognitive a en partie fait dévier cet objectif initial en cherchant dans ces modèles non plus la qualité des prédictions mais l'interprétabilité des données qui sont utilisées pour les réaliser (Hebart and Baker, 2018). Ce domaine a connu ces dernières années un développement très fort qui est dû, parmi de multiple facteurs, au développement dans des versions libres de bibliothèques permettant une démocratisation de ces algorithmes (Pedregosa et al., 2011), ainsi qu'à la mise à disposition de très grandes bases de données ainsi que des ressources computationnelles qui sont souvent nécessaires à leur déploiement. L'intérêt n'est alors pas seulement de pouvoir prédire la modalité de perception d'un stimulus en fonction des patterns d'activation qui sont observés, mais de pouvoir inférer les variables (voxels, électrodes, points dans le temps...) qui participent à encoder ces modalités. Cette approche a d'abord été utilisée dans des paradigmes d'imagerie fonctionnelle en se révélant rapidement plus précise que l'approche univariée pour révéler, au sein même des structures impliquées dans les tâches considérées (e.g. cortex visuel), le détail des cartes d'activations (Quiroga and Panzeri, 2009). La nature même des représentations qu'il est possible de détecter s'en est ainsi trouvée considérablement affinée, permettant de déceler l'orientation (Kamitani and Tong, 2005) des stimuli qui sont présentés au participant ou même de reconstruire les images mentales en l'absence de toute stimulation directe (Lee et al., 2012), voire ses "rêves" durant son sommeil (Horikawa et al., 2013).

Ces méthodes s'intègrent aujourd'hui davantage à la psychiatrie et aux sciences comportementales (Dwyer et al., 2018), à l'électroencéphalographie (Grootswagers et al., 2017; King and Dehaene, 2014), à l'imagerie fonctionnelle ou encore à l'enregistrement de l'activité de groupes de neurones (Wallis, 2018). Appliqués à l'électroencéphalographie, et en particulier à des paradigmes étudiant l'activité évoquée ou induite pendant quelques secondes par la présentation d'un stimulus, il est possible d'utiliser le décodage multivarié pour révéler les patterns spatiaux (i.e. ici les électrodes) qui vont encoder certains traitements cognitifs et d'en observer l'évolution au cours du temps.

Il est par exemple possible de se représenter le niveau d'activation propre à deux conditions enregistrées par l'ensemble des électrodes comprises dans le dispositif (voir Figure 3.1). La stratégie consiste alors à tenter de départager spatialement les niveaux d'activation des deux conditions, par exemple à l'aide d'une droite si nous considérons seulement deux électrodes comme cela est représenté dans la Figure 3.1, ou à l'aide d'un hyperplan si nous considérons l'ensemble des électrodes. Par la suite, la confrontation de ce modèle à de nouvelles données, qui ne figuraient pas dans les mesures utilisées pour l'**apprentissage** de la distribution spatiale, permet d'en évaluer les performances (i.e. les prédictions correctes) dans une phase de **test**. A la différence d'approches traditionnelles basées sur le concept d'**activation** (i.e. la

condition A va davantage activer certaines aires cérébrales), le décodage multivarié est une approche basée sur l'**information** (les informations fournies au modèle permettent ou non de départager les deux conditions) (Hebart and Baker, 2018). Mais ici toutes les différences mesurables entre les conditions d'intérêt, y compris s'il s'agit d'effets non monotones ou non linéaires, sont susceptibles d'être interprétées (Kriegeskorte and Bandettini, 2007). La précision de cette dissociation des patterns d'activité peut alors être évaluée grâce à différentes métriques. Nous utilisons dans la Figure 3.1 et dans l'étude suivante l'aire sous la courbe ROC.

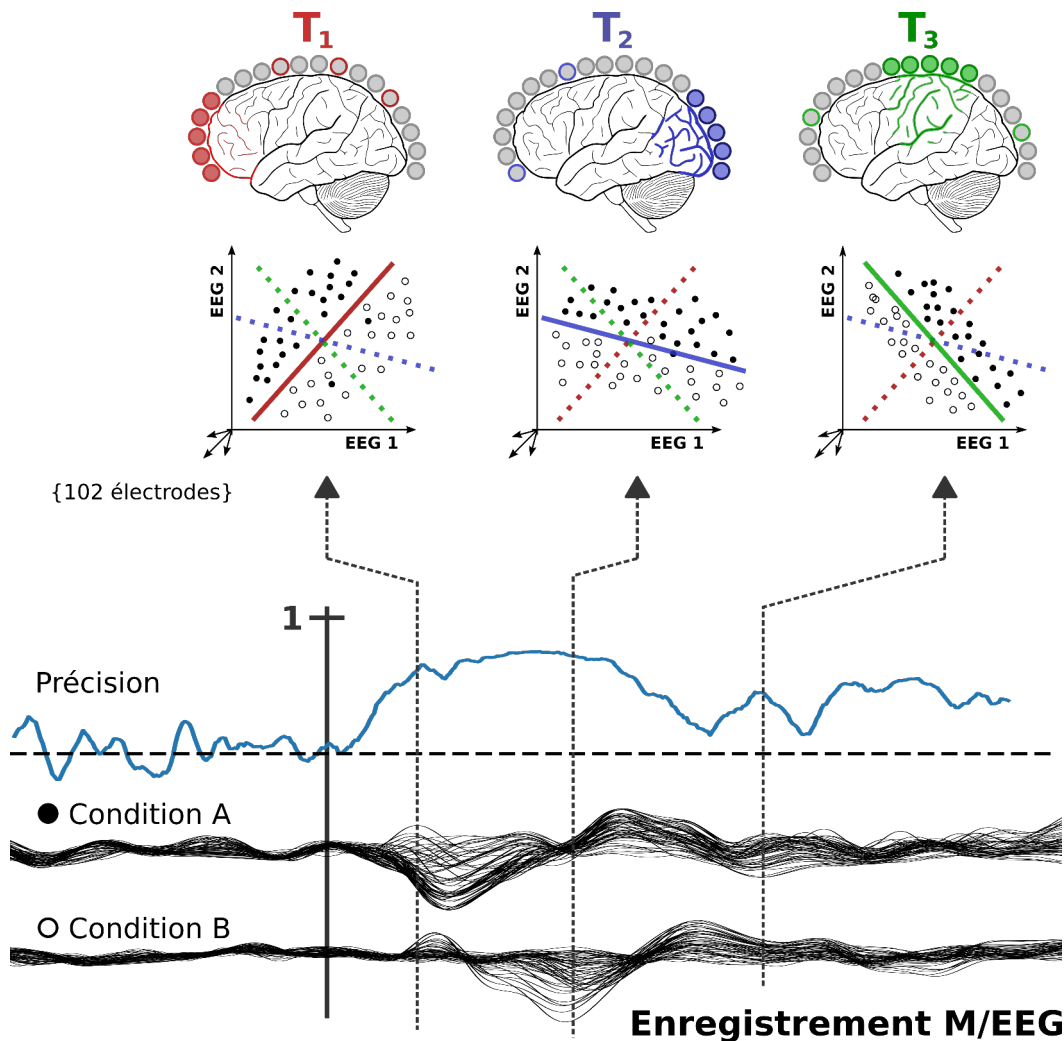


Fig. 3.1: Analyses de patterns multivariés appliquées à l'électroencéphalographie. Pour tout point dans le temps il est possible d'entraîner un modèle pour décodé des informations spécifiques à la tâche (condition, temps de réponse. . .). Les performances de ces modèles peuvent être examinées au cours du temps et vont révéler la dynamique de l'information. Les patterns d'activité sont ici représentés à titre indicatif.

Cette approche permet ainsi de dissocier les patterns d'activité propres à différentes conditions. Toutefois, les processus cognitifs engagés dans le traitement de l'information qu'impliquent certaines tâches peuvent souvent être décomposés

plus finement sous forme d'une séquence contenant des processus distincts. Ainsi que l'illustre la Figure 3.1, la précision de la différentiation des deux conditions ne permet pas immédiatement de visualiser cette succession et de départager les différentes étapes du traitement cognitif qui a été effectué. Pour contourner ce problème, il est possible d'utiliser la méthode de **généralisation temporelle** (King and Dehaene, 2014; „Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method“ 2016; King et al., 2016) qui va utiliser les mêmes modèles que précédemment en cherchant à évaluer leurs performances respectives dans le temps. Dans la méthode précédente, le modèle ayant été entraîné au temps T_1 testait par la suite ses prédictions au temps T_1 uniquement. Mais il est aussi possible d'entraîner un modèle au temps T_1 et d'en évaluer successivement les performances pour tout temps t d'un enregistrement donné. La méthode de généralisation temporelle cherchera alors à savoir si le pattern d'activité qui a été décodé au temps T_1 se retrouve également au temps T_2 (s'il est possible ou non de prédire au-dessus de la chance les conditions en fonction des informations données). La répétition de ce procédé pour tout temps t de l'enregistrement permet alors de visualiser la dynamique d'activation de processus cognitifs différents (voir Figure 3.2).

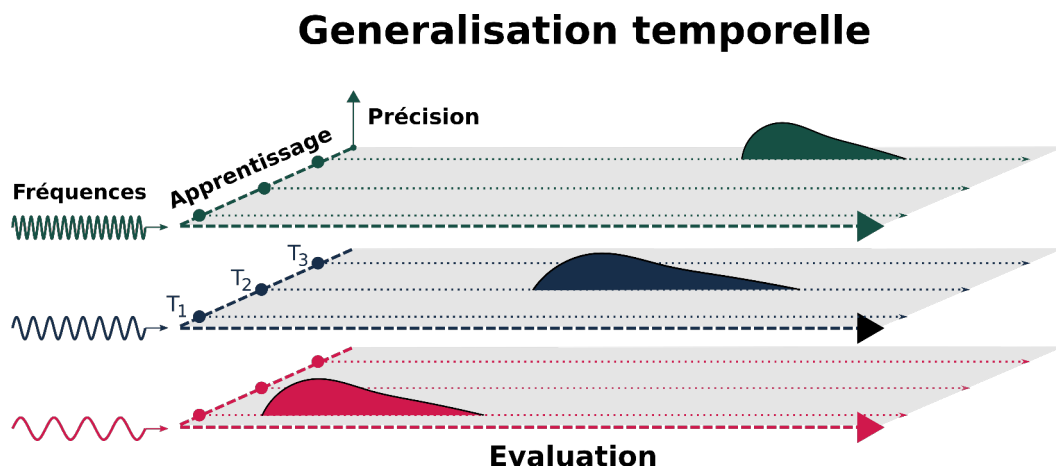


Fig. 3.2: Généralisation temporelle du décodage multivarié appliquée à différentes bandes de fréquences. Il est possible de tester les performances de tout décodeur entraîné sur des instants différents afin de visualiser la réapparition de patterns d'activation similaires. Ici, trois états différents ont pu être dissociés, qui participent chacun à expliquer la différence entre deux conditions mais qui se succèdent de façon séquentielle. D'autres profils, comme la répétition ou le maintien de certains patterns, peuvent être révélés (adapté de (King and Dehaene, 2014)).

Dans l'étude suivante, nous utilisons l'approche de la généralisation temporelle en procédant à une décomposition des bandes de fréquences préalables comme cela est illustré dans la Figure 3.2. Cette étape supplémentaire permet de restreindre les effets que nous observons à l'activité de certaines bandes identifiées au préalable.

3.2 Étude n°2 - Détection multivariée des intrusions mnésiques et étude des processus d'inhibition réactionnels en électroencéphalographie

La seconde étude que nous introduisons ici aborde la question des mécanismes réactifs initiés à la détection des intrusions mnésiques et qui sont considérés jouer un rôle important dans l'inhibition effective de celle-ci. Nous utilisons ici plus en détail les données acquises en électroencéphalographie sur le second groupe de participants de l'étude précédente. Nous avons fait l'hypothèse qu'une intrusion mnésique peut être décodée durant la tâche de "Think/No-Think" car elle va en réalité engager des mécanismes de contrôle inhibiteurs qui sont similaires à ce que déclencherait une intrusion perceptive. En utilisant une tâche attentionnelle originale, nous avons pu extraire, pour chaque participant, le pattern d'activation qui correspond dans son cas à la réorientation de son attention après l'apparition d'un stimulus distractif. En procédant à une généralisation temporelle de ce décodage, nous avons pu estimer la probabilité de sa réapparition lors de la tâche de "Think/No-Think". Cette probabilité nous permettait par la suite d'observer, à chaque essai, la dynamique du réflexe de réorientation attentionnelle et de l'associer au rapport d'intrusion subjectif fourni par le participant. L'examen des activités apparaissant en réaction à ces intrusions nous a permis dans un deuxième temps d'identifier avec plus de précision les mécanismes de l'inhibition mnésique des intrusions. L'article qui est présenté ici a pour titre "*Tracking the emergence and suppression of intrusive memories*".

Tracking the emergence and suppression of intrusive memories

Nicolas Legrand¹, Olivier Etard², Fausto Viader¹, Patrice Clochon¹, Franck Doidy¹, Denis Peschanski³, Francis Eustache¹, and Pierre Gagnepain¹

¹Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France

²CHU de Caen, Service des explorations fonctionnelles du système nerveux, Caen, F-14000, France, Imagerie et Stratégies Thérapeutiques de la Schizophrénie (ISTS), Normandie Univ, UNICAEN, Faculté de médecine, CYCERON, Caen F-14000, France

³European Centre for Sociology and Political Science (Université Paris 1, EHESS and CNRS), Paris, France

Abstract

Pushing unwanted memories out of awareness requires effortful cognitive control. Still, the success of such attempts is prone to a large inter-individual variability and little is known concerning the exact neural sequence supporting proficient inhibitory suppression of unwanted memories. Current theories assume that memory intrusions can be controlled via a reactive prefrontal top-down inhibitory signal to prevent hippocampally triggered neocortical activity. However, direct evidence for such dynamics with the fine-grained time resolution of electrophysiological recordings is still lacking. Here, we used multivariate decoding of EEG recording to track the emergence of intrusive memories during a “Think/No-Think” (TNT) task. Capitalizing on the proximity between perceptual and mnemonic attentional capture, we developed an original attentional task to train a decoder able to detect the presence of perceptual intrusion for each participant. This model was then used to detect similar mnemonic intrusive patterns during the attempts to suppress unwanted memories. The Theta (3-8 Hz) frequency power was associated with the highest decoding performance during the attentional procedure and allowed to decode the emergence of memory intrusions higher than chance during the TNT task. We then analyzed the time-frequency dynamic time-locked to the intrusive memory detection. Results show that intrusion detection was immediately surrounded by an increased Theta activity in the midfrontal and occipital electrodes that predicted later forgetting. These findings support the notion that Theta oscillations are a marker of the successful inhibitory effort to suppress unwanted memories. Critically, they confirm a central hypothesis of neurobiological models of motivated forgetting, which assumes that memory inhibition is triggered in reaction to the competitive interference rising from the counter-intentional retrieval of the intrusive memory.

Keywords: inhibitory control, intrusive memories, EEG, Theta, multivariate decoding, generalization across time, random forests

Introduction

A fundamental dynamic of motivated forgetting involves the intrusion of unwelcome content, coupled with the goal of excluding that content from consciousness. Controlling intrusive memories, thoughts, or emotions, is a challenging task that is prone to a large inter-individual variability (Levy & Anderson, 2008). Neurobiological models of motivated forgetting (Anderson & Huddleston, 2011; Anderson, Bunce, & Barbas, 2016) assume that inhibitory control of memory awareness is performed by a reactive process to prevent hippocampally triggered neocortical activity (Anderson, 2003; Hu, Bergström, Gagnepain, & Anderson, 2017). Inhibitory control is achieved by the right dorso-lateral prefrontal cortex (Anderson, 2004; Depue, Curran, & Banich, 2007) which applies a down-regulation on memory-related areas (Gagnepain, Henson, & Anderson, 2014; Gagnepain, Hulbert, & Anderson, 2017; Benoit & Anderson,

2012; Benoit, Hulbert, Huddleston, & Anderson, 2015). In these studies, the negative top-down inhibitory signal increases when counter-intentional intrusion emerges into a person's awareness (Gagnepain, Hulbert, & Anderson, 2017; Benoit, Hulbert, Huddleston, & Anderson, 2015). This view is supported by several fMRI findings, also evidencing the extension of this targeted suppression to the amygdala (Gagnepain, Hulbert, & Anderson, 2017) and the visual cortex (Gagnepain, Henson, & Anderson, 2014), presumably making suppressed memories less vivid and distressing over time. However, while pervasive impairments of inhibitory control is a robust correlate of psychiatric disorders (Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015), the control neural mechanisms shielding against intrusions and inhibiting memories remain unclear, and the reactive component of inhibitory control central to neurobiological accounts of motivated forgetting remains yet to be directly evidenced.

Previous works made an intensive use of “Think/No-Think” (TNT) paradigm (Anderson & Green, 2001) as an experimental procedure to induce intrusive memories. However, evidence supporting the proposed cognitive and neural dynamics of intrusion purging is still indirect. The existing literature grounding on the high temporal resolution of MEEG recordings mainly used event-related potential (ERP) (Bergström, Anderson, Buda, Simons, & Richardson-Klavehn, 2013; Bergström, Velmans, de Fockert, & Richardson-Klavehn, 2007; Chen et al., 2012; Hanslmayr, Leipold, Pastotter, & Bauml, 2009; Hellerstedt, Johansson, & Anderson, 2016; Mecklinger, Parra, & Waldhauser, 2009; Bergström, de Fockert, & Richardson-Klavehn, 2009; Bergström, de Fockert, & Richardson-Klavehn, 2009) or time-frequency analysis (Ketz, O'Reilly, & Curran, 2014; Waldhauser, Bäuml, & Hanslmayr, 2014; Depue et al., 2013; Waldhauser et al., 2018; Legrand et al., 2018). These approaches analyze the regulation of competing mnemonic interference time-locked to the onset of the reminder and average large periods of presumed intrusions, leaving uncertain that the observed neural response to regulate memory interference is truly a reactive inhibitory process. Given the low temporal resolution of fMRI acquisition, the same problem arises with previous fMRI studies evidencing a negative coupling between frontal and medial temporal lobe structures as a marker of intrusion control (Gagnepain, Hulbert, & Anderson, 2017; Benoit, Hulbert, Huddleston, & Anderson, 2015). Although there is a high unpredictability of memory intrusion within a trial, expectation signal derived and updated from the temporal structure of the intrusive patterns across the task could still. Previous fMRI evidence interpreted as reactive inhibition could, therefore, be confounded with attempts to proactively control an expectation signal increasing before an intrusion. Here, we attempt to tackle this problem

using multivariate decoding of EEG recording to track the emergence of intrusive memories during a TNT task and assess the oscillatory dynamics time-locked to the emergence of intrusions into a person's awareness.

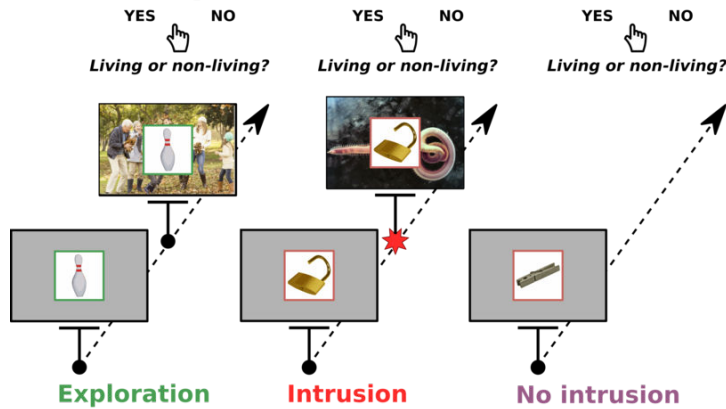
Detecting intrusion from noisy electrophysiological signal can be challenging. One could address this problem by refining the hypothesis concerning the spatial or frequency domain of interest from priors studies, which would require strong evidence concerning the actual properties of memory intrusions. Here, we detected intrusive memories during the TNT task in time, space and frequency domains for each trial and each participant using an original attentional model for which the apparition of a perceptual intrusion was controlled. Memory intrusions take the form of counter-intentional and unwanted abrupt onset of memory trace in awareness, that requires inhibitory control to be purged and ignored (Levy & Anderson, 2012). The neural system underlying the attentional capture of abruptly onsetting stimuli in the perceptual domains may play an important role in the neurobiological model of memory intrusion. Attention to memory or perceptual content is subserved by shared neural structures (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Cabeza et al., 2011; Ciaramelli, Grady, Levine, Ween, & Moscovitch, 2010; Ciaramelli, Grady, & Moscovitch, 2008). It may be hypothesized, therefore, that counter-intentional retrieval of an unwanted memory reflects involuntary mnemonic capture driven by reflexive orienting mechanisms similar to other perceptual orienting mechanisms. We thus designed an original attentional procedure reproducing the involuntary capture elicited by irrelevant sensory information during the TNT task.

Attention and cognitive control have been found to modulate low-frequency EEG activity (< 30 Hz) (Cavanagh & Frank, 2014; Helfrich & Knight, 2016; Klimesch, 2012). Theta activity (3-8 Hz) is notably associated with binding and synchrony of distant region in the brain (Clouter, Shapiro, & Hanslmayr, 2017; Zhang, Watrous, Patel, & Jacobs, 2018). When recorded in the midline frontal electrodes, theta power is thought to be dominated by the activity in the medial prefrontal cortex (Hanslmayr et al., 2008; Womelsdorf, Johnston, Vinck, & Everling, 2010) and may coordinate activity in the frontal regions. An increase of frontal midline has been related to attention orientation (Rajan et al., 2018; Kam, Solbakk, Endestad, Meling, & Knight, 2018), cognitive control mechanisms (Popov et al., 2018; Oehrns et al., 2018; Mückschel, Dippel, & Beste, 2017; Bramson, Jensen, Toni, & Roelofs, 2018) and memory interference detection (Ferreira, Maraver, Hanslmayr, & Bajo, 2018; Staudigl,

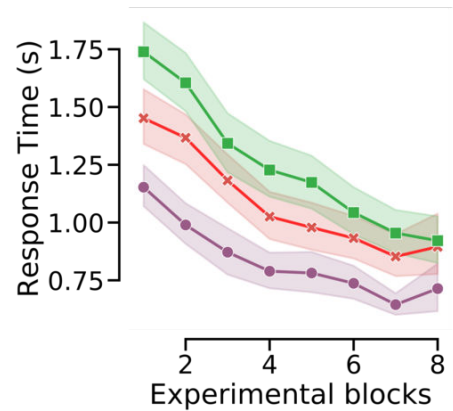
[Hanslmayr, & Bauml, 2010](#); [Töllner et al., 2017](#); [Hanslmayr et al., 2012](#); [Hanslmayr, Staudigl, Aslan, & Bäuml, 2010](#)).

These considerations led us to hypothesize the existence of an inhibitory control process reflected by frontal theta oscillations that are engaged to purge awareness from unwanted content and suppress on the long-term the unwanted expression of memory. To test this hypothesis, participants underwent an attentional task before the realization of the TNT task. The purpose of this procedure was to create a situation in which the presentation of a distracting visual stimuli interferes with goal-oriented processes initiated by the participant, a process identical to the interfering competition arising from intrusive memory in the TNT task and expected to trigger a top-down inhibition to reorient the focus of attention on the targeted stimulus and away from the interfering signal. Intrusions often take the form of visual recollection of episodic details, whose EEG correlated are often localized in occipital sensors ([Wimber, Maaß, Staudigl, Richardson-Klavehn, & Hanslmayr, 2012](#)). We hypothesized that we could decode the occurrence of such intrusive visual patterns during the TNT task and therefore the precise moment when intrusive memory enters awareness to track down the reactive oscillatory dynamics during attempts to suppress and regulate those intrusions.

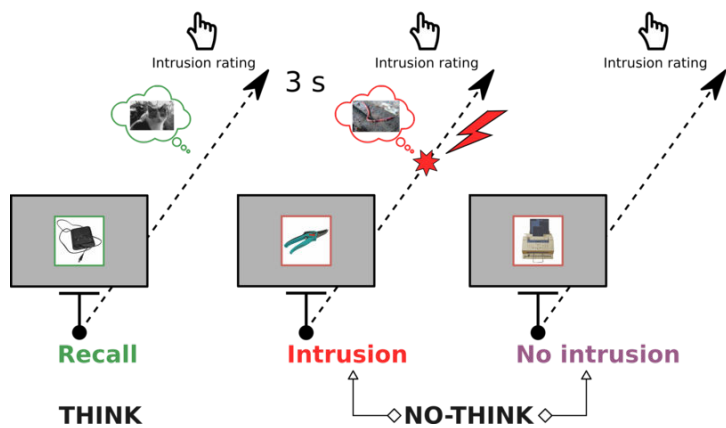
a. Perceptual intrusion



Attentional task performances



b. Memory intrusion



Post-TNT memory performances

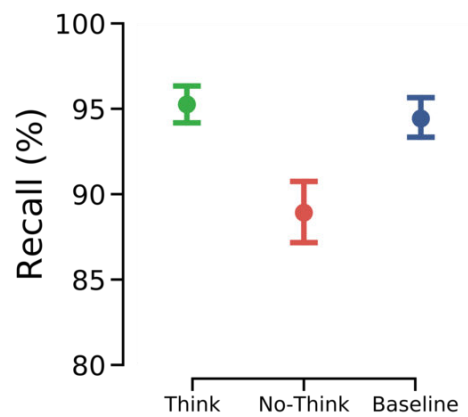


Figure 1: Experimental design and behavioral results.

Legend figure1: A. [left panel] During the attentional procedure, the participant had to categorize visual stimuli as fast as possible. Each trial started with a picture of an object presented at the center of the screen surrounded by either a green or a red box. For green trials, a larger image was presented in the background, 200 milliseconds following the initial object presentation. The participant had the instruction to explore this image and to indicate whether humans were present in the picture or not. For red trials, participants had the instruction to stay focused on the image on the screen and had to indicate whether an object or an animal was represented. On half of the trials, an image appeared in the background (Intrusive condition). As for No-Think trials during the TNT task, participants were instructed

to try their best to block the images from entering their consciousness and to maintain their attention on the central object [**right panel**] Response time associated with Intrusion, Non-Intrusion and Exploration conditions. Overall, responses speed increases across the eight experimental blocks. However, both Intrusion and Exploration conditions were associated with longer response time during the overall experiment, while this difference tended to be reduced with the repeated attempts to block intrusive images across blocks. **B.** [**left panel**] “Think/No-Think” procedure. After the initial encoding of object-scenes pairs, Think and No-Think items were presented with green or red boxes respectively. For green (Think) items, participants were told to generate as detailed and complete an image of the associated scene as possible. For red (No-Think) items, the cue was bounded by a red box, and participants were told that it was imperative to blank their mind and to prevent the scene from coming to mind at all and that they should fixate and concentrate on the object-cue without looking away. If the object image came to mind anyway, they were asked to push it out of mind. During red-cued trials, participants were asked to block thoughts of the scene by blanking their mind and not by replacing the scene 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 offset of each of the Think or No-Think trial cues, participants reported the extent to which the associated scene had entered awareness. [**right panel**] Recall of the associated scenes after the TNT procedure. Memory performances for No-Think items were decreased compared to both Think and Baseline items.

Results

Behavioral

Validation of the attentional model of intrusion - Intrusive images interfere with attentional focus. First, we wanted to confirm that the Intrusion condition during the attentional procedure induced the expected attentional capture. We hypothesized that this phenomenon would increase the response time due to the necessity for the participant to inhibit the irrelevant stimulus and reorient his/her focus before performing the classification. We averaged the response time for each of the attentional conditions (Explore, Intrusion, Non-Intrusion) and each of the eight experimental blocks (see Fig. 1 a.). We only report here the trials that survived the artifact correction during EEG preprocessing (see). A

Block x Condition ANOVA on the averaged response times revealed a significant effect of Condition ($F_{(2,56)} = 77,61, \eta_g^2 = 0.73, p < 0.001$), Blocks ($F_{(7,196)} = 83,25, \eta_g^2 = 0.75, p < 0.001$), and an interaction between these two factors ($F_{(14,392)} = 8,61, \eta_g^2 = 0.23, p < 0.001$). Post-hoc comparisons with Bonferroni correction revealed that response time were shorter for the Non-intrusion condition, both compared to the Intrusion ($t_{(26)} = -10.24, d = -1.9, p < 0.001$) and Exploration ($t_{(26)} = -10.67, d = -2.0, p < 0.001$) conditions. This longer response time under intrusive condition compared to a situation without any perceptual interference suggest that the abrupt onset of an interfering background image has induced a reflexive orienting process slowing down the reaction time to the attended central object. We interpreted this result as a corroboration of our perceptual intrusion model.

Aftereffect of inhibitory control on memory performances. Then, we tested the presence of a successful inhibitory control effect by comparing the subsequent memory performances on Think, No-Think, and Baseline conditions. A repeated measure ANOVA on the number of correct recall in the final test with Condition as factor revealed a significant difference between Think, No-Think and Baseline conditions ($F_{(2,52)} = 12,70, \eta_g^2 = 0.33, p < 0.0001$). We controlled for the presence of false recalls via the audio recording of the participant's description of the recalled scene. Baseline items had been trained in the same initial learning session with the Think and No-Think items but were omitted from the Think/No-Think phase. Because these items were trained at the same time as the Think and No-Think items but did not participate in the Think/No-Think task, they provide a baseline estimate of memorability of the scenes, given that neither suppression nor retrieval had been performed on them. These pairs enabled us to assess the effects of retrieval and suppression on the retention of Think and No-Think items respectively, against pairs that were similarly old. Critically, while no difference was found between Think and Baseline items ($t_{(26)} = 0.49, p = 0.37, d = 0.12$), the planned comparison between No-Think and Baseline items revealed that participants recalled significantly fewer items under the No-Think condition ($t_{(26)} = -3.25, p = 0.003, d = -0.64$). This results then confirmed that the "Think/No-Think" procedure induced a successful inhibition of unwanted memories.

EEG

Multivariate decoding

Decoding the interference of intrusive image during the attentional procedure. Grounding on the confirmed hypothesis that the Intrusion presentation, 200 ms after the onset of the object, interfered with the ongoing classification task as reflected by the longer response time, we wanted to investigate the oscillatory dynamics supporting this attentional capture. Several recent works pointed to the Theta (3-8 Hz) oscillatory power as a robust correlate of stimulus coding and interference detection-resolution via cognitive control (Mückschel, Dippel, & Beste, 2017; Staudigl, Hanslmayr, & Bauml, 2010; Hanslmayr, Staudigl, Aslan, & Bäuml, 2010; Cohen, 2014). We used the power of this Theta frequency band as recorded in our 102 electrodes of interest in the subsequent analysis. To explore the temporal dynamics of interference detection, we decoded the condition (Intrusion vs Non-intrusion) over time by testing and training a random forest classifier at each time points between -500 to 1500 ms around the first stimulus presentation (see Method). These classifiers were embedded in a 10 folds cross-validation framework, and participant performances were measured through the averaged Area Under the Curve (AUC) metric. Results are presented in Figure 2 a. We then tested at each time point the difference from chance ($AUC = 0.5$) using a cluster-based permutation t-test (Maris & Oostenveld, 2007) as implemented in the *permutation_cluster_1samp_test()* function of the MNE Python package (Gramfort, 2013). We found a first significant decoding window between 200 ms and 600 ms after the trial start (cluster p -value=0.0002) with the maximal decoding accuracy ($AUC = 0.65$; $SD = 0.08$); arising 140 ms after the presentation of the intrusive image. A late significant window between 620 ms to 800 ms was also found but the associated decoding accuracy was modest (cluster p -value=0.02). This confirmed the role of Theta activity to detect the emergence of intrusive and distracting visual stimuli. To corroborate this idea, we also explored the spatial distribution of Theta power and contrasted Intrusion and Non-Intrusion condition for each participant. For each electrode, we averaged the signal over bins of 100 ms on a window of interest ranging from 200 ms to 600 ms after the stimulus presentation. We restricted our analysis to this interval as it overlapped with the first significant temporal cluster of decoding. We then used a one-sample T-test (*permutation_cluster_1samp_test* function) to reveal any difference between Intrusion and Non-Intrusion, and used the connectivity matrix of the EEG montage to inform the spatial dependence of our measurements and adjust the family-wise error rate. The resulting t-values and significant cluster are reported in the lower

panel of Figure 2. Occipital electrodes were associated with significantly higher Theta power during Intrusion versus Non-Intrusion. This pattern also spreads to the midline frontal electrodes 200 ms after the intrusion presentation. Interestingly, this time-window overlaps with the peak of the performance of the intrusion decoding classifier. This suggests that this pattern of Theta power increase is a marker of intrusion detection and resolution. Moreover, to confirm that this optimal decoding window is mainly driven by the Theta activity, we also repeated this decoding analysis for each frequency bin from 3 to 30 Hz. Results confirmed the higher decoding accuracy of the Theta band with a peak at 5 Hz (see *Supplementary materials*, Figure S1).

Decoding the intrusion of unwanted memories during the “Think/No-Think” task based on the attentional model of interference. Based on the previous results, we wanted to generalize the decoder to the Think/No-Think task. This approach corresponds to the temporal generalization method (King & Dehaene, 2014; Dehaene & King, 2016), for which decoder trained at one time point are tested on another time point to infer the dynamic of mental representation. Here, we trained another random forest classifiers using all the trials from the attentional task and tested their prediction to decode Intrusive versus Non-intrusive No-Think trials based on the introspective report of the participant. This analysis was restricted to the decoders trained and time-locked to immediate 200 ms time-window following the interfering image presentation. This time-window included early intrusive components successfully decoded, up to the maximal attentional decoding accuracy (i.e; 140 ms post-intrusion) but stopped at 200 ms post-intrusion when the decoder performances drop drastically (Figure 2 a). Moreover, our goal was to explore the temporal dynamics of intrusive patterns over the 3 seconds following the onset of the No-Think trials. So instead of predicting the category and estimating the AUC at each time point, we estimated the probability associated with an Intrusion pattern for the classifier time-locked to perceptual intrusion (Figure 2 b). We did so because a memory intrusion can appear at any moment over the whole duration of the No-Think trials. Thus, computing prediction performances at each time point may inaccurately reflect our decoders precision. Then, we tried to predict memory intrusion as accurately as possible from this decoding probability time-course. Considering that a memory intrusion appears abruptly, we labeled as intrusion the peaks in decoding probability. However, given that intrusion decoding probability varies widely across items and participants, we varied the detection threshold of the intrusion probability peak (ranging from 0.5 to 1 with 0.02 incremental steps) to optimize decoding performances. This method allowed us to combine an optimal temporal classifier and de-

tection threshold that most accurately reflected the introspective report provided by the participant (Intrusion or Non-Intrusion). The resulting patterns decoded memory intrusions higher than chance (mean AUC: 0.60, AUC standard deviation: 0.08; $t_{(26)}=6.83$, $p<0.0001$). The results are displayed in Figure 2 c. In Figure 2 d we show the averaged best decoding pattern among the group of participants. This corresponds to the contrast between the Intrusive and Non-Intrusive conditions from the attentional task. A cluster of significant occipital electrodes was found. These findings confirmed that memory intrusions were best predicted by early visual processing associated with perceptual interference detection.

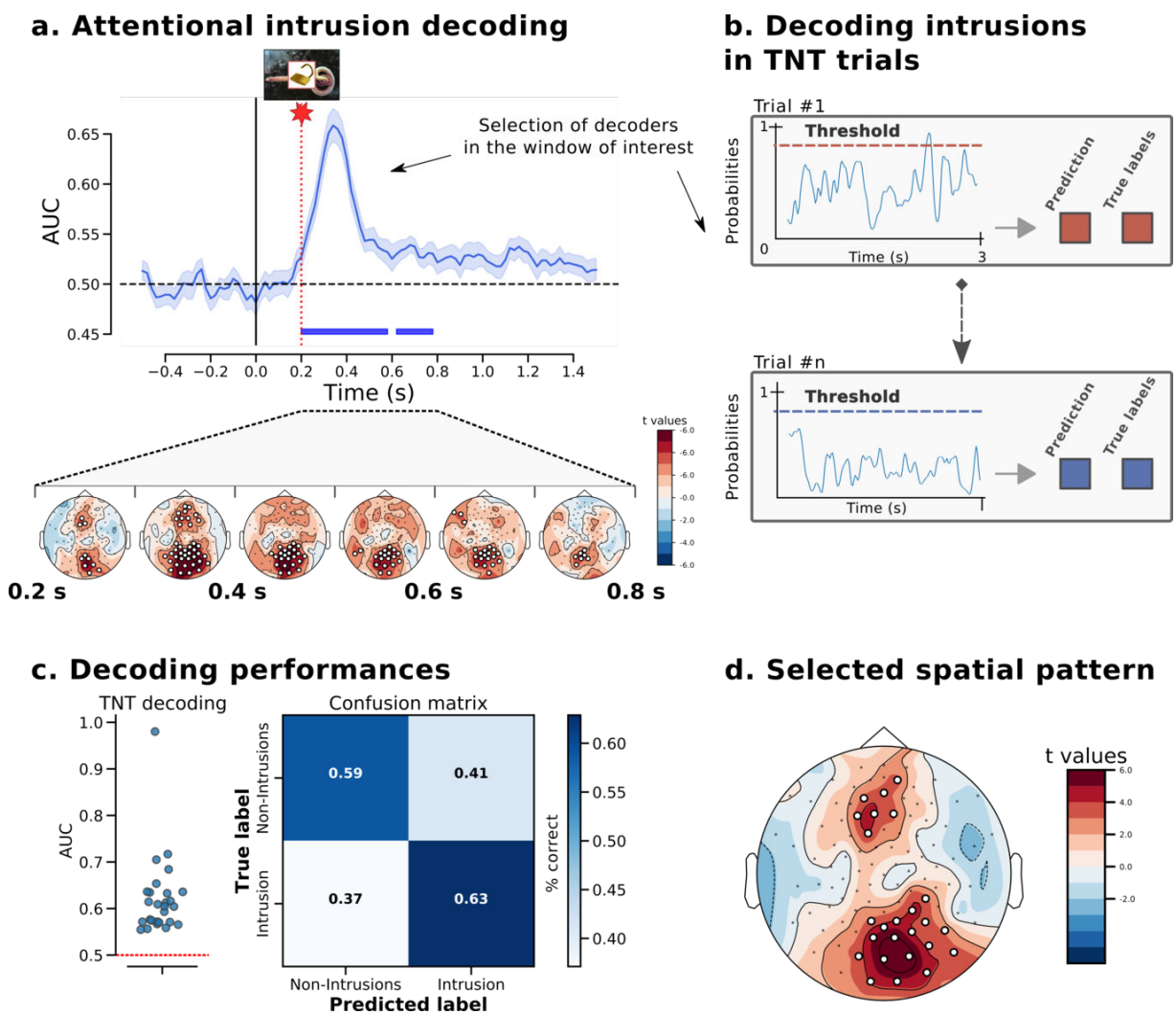


Figure 2: Decoding performances during the attentional and TNT procedures.

Legend figure 2: **A.** *Random forest classifiers showed performances significantly higher than chance from 220 ms to 540 ms, as well from 639 ms to 800 ms, following the onset of the initial object presentation (intrusive image appears at 200 ms and is marked with a red dashed line). The topographic maps represent the contrast between Intrusion and No intrusions conditions. Higher decoding scores were associated with a significant Theta (3-8 Hz) increase in a set of occipital and frontal midline electrodes.* **B.** *We selected classifiers in the 200 ms time-window following the onset of the intrusive image presentation in order to include the maximal attentional decoding accuracy and to capture early visual interference processes. Then, the classifiers trained on the attentional task were used to predict the probability of an Intrusion condition during the No-Think trials from the “Think/No-Think” procedure.* **C.** *The resulting AUC and the confusion matrix show that the resulting predictions were significantly higher than chance.* **D.** *The statistical spatial pattern from the attentional procedure associated with the higher prediction values during the TNT, when contrasting perceptual Intrusion and Non-intrusion.*

Time-frequency changes time-locked to decoded intrusions during the TNT procedure

Next, we thought to explore the oscillatory dynamics surrounding the occurrence of memory intrusions. Based on the precision of selected intrusion decoding predictions, we locked the onset of the time-frequency analysis (ranging from -500 ms to 800 ms) to the maximal intrusion probability identified in the suppression time-window (excluding decoded intrusion at the very end of the trial to avoid any confusion and overlap between the time-frequency analysis and subsequent introspective judgment). Figure S2 in the **Supplementary material** reports the temporal distribution of decoded intrusions used for time-locking the time-frequency analysis. Here, we only analyzed trials labeled as intrusive both by our classifiers and by the participant. Based on the observed distribution, we excluded intrusions detected during the 500 ms following the presentation of the image to avoid overlap with evoked potential. For the remaining intrusions detected, we extracted the complete time-frequency representation of all No-Think items (see **Method** section for details). However, in order to compare the time-frequency dynamic following an intrusion with non-intrusive recording, we used the available Non-intrusion condition as labeled by the participant and extracted time-frequency representation using the same method. Finally, we applied a baseline correction by subtracting the mean of the segment from -500 ms to -200 ms before the

intrusion, averaged and contrasted the Intrusion and Non-intrusion condition.

We explored the temporal dynamics of power following the onset of decoded memory intrusion across three frequency bands of interest: Theta (3-8 Hz), Alpha (8-13 Hz) and Beta (13-30 Hz). These frequencies are thought to play a key role in the organization of complex cognitive processes, but they can also exhibit divergent relations with the underlying brain activity. For example, Alpha, and to a lesser extent Beta power, can be negatively correlated with the BOLD signal response of the spiking activity in the region of interest (Hanslmayr et al., 2011). For these reasons, and also because we had a strong hypothesis concerning the role of midline theta power in inhibitory control under intrusive conditions (Cavanagh & Frank, 2014; Hanslmayr et al., 2008; Ferreira, Maraver, Hanslmayr, & Bajo, 2018; Hanslmayr, Staudigl, Aslan, & Bäuml, 2010), we examined separately the temporal dynamics of these three frequency bands. We averaged the log-ratio of the frequency power (to normalize frequency changes across selected bands) across bins of 200 ms from -200 ms to 800 ms time-locked to the detected memory intrusion and contrasted with the Non-Intrusion condition.

Results from the cluster-based permutation test (see Figure 3 a.) showed a strong transient increase of Theta activity in both occipital and midline frontal electrodes from -200ms before intrusion to 200 ms after intrusion (all cluster p -values < 0.001). This increase in Theta activity surrounding the memory intrusion was not sustained and disappears 200 ms after the detection of the memory intrusion. This result is consistent with the apparition of a conflict-resolution process mediated by Theta oscillations in reaction to interfering mnemonic activity in posterior areas. We did not find further significant effects expressed in cluster larger than 3 electrodes in the others frequency bands.

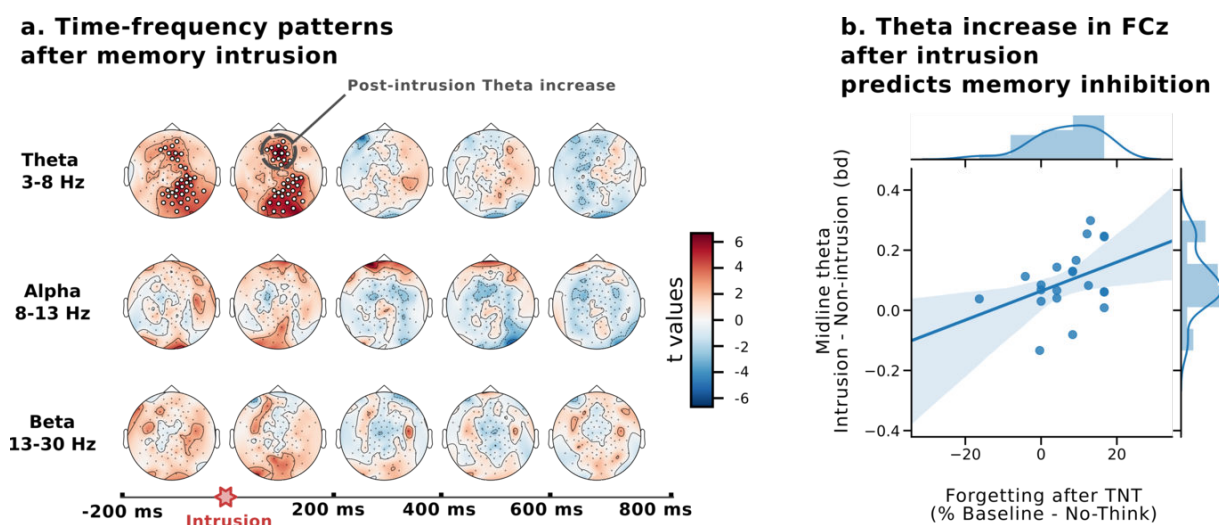


Figure 3: Post-intrusion activation patterns for Theta, Alpha and Beta frequency bands.

Legend figure 3: a. For each participant ($n=21$), we used the decoders prediction to time-lock time-frequency dynamics to the onset of intrusion detection ($t=0$). Power was averaged over time-slice of 200 ms. The color range represents the t values from the one-sample permutation t -test contrasting Intrusion and Non-Intrusion conditions. Clusters of electrodes where values significantly differed from 0 are represented by white circles. **b.** Correlation between Theta increase in midline frontal electrodes during memory intrusion and later forgetting.

According to neurobiological models of motivated forgetting, memory inhibition is triggered in reaction to the competitive interference rising from the involuntary retrieval of the intrusive memory. Whether inhibition disrupts memory traces and produces a suppression effect, would be expected to depend, therefore, on how effectively frontally-mediated mechanisms detect and then mitigates unwanted intrusions through inhibitory control. To test this possibility, we examined the relationship between the Theta power increase in midline frontal cluster time-locked to the intrusion detection, and memory suppression scores (Baseline - No-Think). The difference in final recall between the items associated with the No-think condition and the items associated with the Baseline condition (that were not presented during the TNT procedure) is an index of inhibitory-induced forgetting. We restricted this

analysis to a -100/0 ms and 0/100 ms time-windows time-locked to the intrusion detection and averaged the power across time-points for the frontal midline electrode that was associated with the maximal Theta increase. As predicted, results revealed a positive correlation between Theta increase and memory inhibition immediately after the intrusion ($r=0.38$, $p=0.04$, one-tailed pearson correlation; Figure 3 b.), but not immediately before ($r=0.18$, $p=0.20$, one-tailed pearson correlation), suggesting that the ability to inhibit unwanted memory traces is closely linked to their initial unintentional and involuntary reinstatement.

Discussion

In this work, we investigated the oscillatory sequence supporting the reactive inhibitory control of intrusive memory during a “Think/No-Think” (TNT) paradigm. Using supervised machine learning approach, we first trained a random forest classifier to detect early sensory interference during the emergence of perceptual distraction, simulating the unwanted emergence of a mnemonic intrusion into a person’s awareness. This hypothesis is motivated by the fact that counter-intentional retrieval of an unwanted memory, might reflect involuntary attentional capture driven by reflexive orienting mechanisms similar to other perceptual orienting mechanisms (Cabeza et al., 2011; Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Chun & Johnson, 2011). This attentional model of intrusion was then used to infer the occurrence of similar intrusive patterns when participants had to control memories during the “Think/No-Think” procedure. Our results confirmed the close relationship existing between perceptual and mnemonic capture, and revealed the prominent role of Theta (3-8 Hz) activity, mainly located over occipital sensors, to decode intrusive reflexive orienting toward mnemonic input from a perceptual interference training model. Time-locking the oscillatory dynamics to the onset of decoded intrusion additionally revealed a frontal Theta increase preferentially located over midline sensors. Critically, these anterior oscillatory dynamics time-locked to intrusive memory, and presumably reflecting the extensively described fronto-parietal network engaged during inhibitory control (Gagnepain, Henson, & Anderson, 2014; Gagnepain, Hulbert, & Anderson, 2017; Benoit & Anderson, 2012; Benoit, Hulbert, Huddleston, & Anderson, 2015), predicted later forgetting. These findings thus confirm a central hypothesis of the neurobiological model of motivated forgetting, which assumes that memory inhibition is triggered in reaction to the competitive interference rising from the involuntary retrieval of the intrusive memory.

Decoding perceptual attention to intrusion. Multivariate decoding applied to the sensor space during the attentional task revealed that the emergence of perceptual distracting images outside the focus of attention could be accurately decoded from 20 ms to 340 ms following the onset of the perceptual intrusion. Behavioral data indicated that an interference effect, consisting in a slower processing time of the central object, parallel the emergence of the background intrusive images. Taking together, these findings suggest that the abrupt onset of a distracting image triggered a reflexive orienting response toward the interfering image whose process compete with the processing of the central attended object. Such competition between the semantic processing of the central object and the interfering image might arise from an early visual categorization process (< 200 ms) of intrusive images, consistent with several other ERP and MVPA studies (Simanova, van Gerven, Oostenveld, & Hagoort, 2010; Robinson et al., 2017). Our findings thus suggest that perceptual intrusions are associated with increased early sensory processing that quickly reaches perceptual awareness (Dijkstra, Mostert, de Lange, Bosch, & van Gerven, 2018; King, Pescetelli, & Dehaene, 2016). Critically, however, the models trained on the time-point preceding the intrusive image presentation were not associated with a similar increase, which suggests that the random forest captured discriminant features in this interval to differentiate between Intrusive and Non-intrusive trials. The absence of significant decoding performances between 0 and 200 ms also highlight that decoding performances cannot be confounded by the event-related potential elicited by the visual presentation of the pre-intrusion object on the screen. Low performances before the interfering image presentation thus confirm the idea that our models learned to contrast patterns associated with simple visual and decision processing with the exact same patterns plus an interfering visual stimulus.

Concomitantly, we observed a significant increase of Theta power in a spatial pattern comprising both occipital and midline frontal electrodes clusters during the during competitive interference of intrusion condition, compared with the Non-Intrusion condition that does not involve such a competition. The strength of this activation differences was maximum during the peak of the classifiers performances during competitive interference. Although we did not localize the source of these scalp differences, their location is consistent with the presence of an early visual component that might be associated with an interference detection and mediation process implemented by the prefrontal midline structure (e.g. (Staudigl, Hanslmayr, & Bauml, 2010)). Taken together, these data provide additional support to the notion that Theta power is engaged in interference detection and resolution (Sauseng et al.,

2005; Hanslmayr et al., 2008; Cooper et al., 2015).

Occipital activation patterns in the Theta frequency band successfully decode the intrusion of unwanted memories during memory control. We then hypothesized that the early sensory reflexive orienting response toward perceptual intrusion could serve as a template to track the emergence of mnemonic intrusion into awareness during the TNT task. Analysis revealed that decoding accuracy of Intrusion vs. Non-Intrusion during the TNT task was significantly above chance. Multivariate decoding is increasingly used in memory research as a way to track the reactivation of previously learned material in high dimensional physiological recording and has been successfully applied under a variety of conditions (Kerrén, Linde-Domingo, Hanslmayr, & Wimber, 2018; Schönauer et al., 2017; Schreiner, Doeller, Jensen, Rasch, & Staudigl, 2018). These previous works highlight a fundamental role of Theta oscillation to the signature of memory reactivation. Although we are not tracking here the memory reinstatement of a particular stimulus, our findings also support a core role of Theta oscillations when a mnemonic input is reflected into awareness. In addition, our findings further demonstrate a general role of Theta oscillations to encode an abrupt onset in the sensory domains indistinctively from perceptual (external) or mnemonic (internal) input. Thus, attentional mechanisms reflecting interfering sensory inputs play an important role in the neurobiological model of memory intrusion. The unwanted emergence of a memory into a person's awareness is driven by reflexive orienting mechanisms similar to other perceptual orienting mechanisms. This is in line with findings highlighting the shared neural structures that subserve attention to memory and perceptual content (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Cabeza et al., 2011; Ciaramelli, Grady, Levine, Ween, & Moscovitch, 2010; Ciaramelli, Grady, & Moscovitch, 2008; Chun & Johnson, 2011).

Midline frontal Theta increases during memory intrusion and predicts memory inhibition. The detection of a memory intrusion further revealed a significant increase of Theta power in the frontal midline electrodes. This increase was specifically found between -200 and 200 ms around the peak of intrusion decoding. Theta power in the midline frontal electrodes is a robust marker of cognitive control (Eschmann, Bader, & Mecklinger, 2018; Cavanagh & Frank, 2014). This frequency is central both for the orchestration of frontal dynamics (Cavanagh, Zambrano-Vazquez, & Allen, 2012; Hauser et al., 2014) but also for the synchronization of information flow between distant brain areas, to flexibly distinguish relevant and irrelevant information (Verguts, 2017) or mobilize executive functions (Mizuhara &

Yamaguchi, 2007). Critically, this increase in Theta activity after the detection of intrusive memory was furthermore predictive of later forgetting when comparing No-think and Baseline conditions. This suggests that this reactive activity might not only reflect the detection of memory reactivation and its interference, but also the voluntary effort to push the memory out of consciousness and to regulate its competitive interference, which is closely related to the ability to inhibit the suppressed memory traces. This is well in line with previous works linking Theta midline activity to the regulation and mediation of interfering memory signal competing with selected memory traces in the context of the retrieval-induced forgetting paradigm (Hanslmayr et al., 2008; Staudigl, Hanslmayr, & Bauml, 2010), or with the ability to suppress memories in the context of TNT paradigm (Depue et al., 2013).

These previous works, however, time-locked their analysis to the onset of the stimulus triggering the competing mnemonic interference and averaged large periods of presumed intrusion or interference across trials, leaving uncertain that the observed neural response to regulate memory interference is truly a reactive inhibitory process. Given the low temporal resolution of fMRI acquisition, the same problem arises with previous fMRI studies evidencing a negative coupling between frontal and medial temporal structures as a marker of intrusion control (Gagnepain, Hulbert, & Anderson, 2017; Benoit, Hulbert, Huddleston, & Anderson, 2015). Indeed, contrasting interfering and non-interfering situations (i.e. Intrusion versus Non-Intrusion in the context of TNT paradigm) does not strictly rule out alternative explanations unrelated to a reactive inhibitory control process. For instance, according to the minimization of the free-energy principle (Friston, 2009), it is very likely that participants, through the course of the TNT task, learn an internal model generating beliefs about the expected probability of intrusive memories, either through the course of the task or bound to a particular trial. Thus, it might be the case that previous evidence suggesting that inhibitory control is engaged reactively to regulate intrusive interference reflects in fact attempts to control an expectation signal (which is presumably higher during intrusion than non-intrusion trial, thereby confounding this difference). This issue is greatly mitigated if not absent with our decoding approach because decoded intrusions arise at any time of the suppression trials (see Supplementary material, Figure S2). This great variability of abrupt intrusive onsets makes it very unlikely that this intrusive signal is coupled to an expectation signal derived from the sequential structure of intrusions across the task and unavoidably time-locked to the appearance of the object-cue triggering such expectations.

Moreover, the intrusion introspective report remained dependent on individual and subjective decision criterion threshold. Thus, for a similar intrusive sensory signal, a participant can interpret this signal as intrusive while another not. This was more rigorously controlled in this study as intrusions were detected based on a mathematical model (random forests classifiers) using a threshold that may vary across participants. Thus, although previous fMRI findings (Gagnepain, Hulbert, & Anderson, 2017; Benoit, Hulbert, Huddleston, & Anderson, 2015) establishing the presence of a negative inhibitory signal via connectivity measures could reflect a complex mix between proactive, reactive inhibitory and subjective individual decision criterion, our decoding approach of intrusive memory combined with the time resolution of the EEG recording allowed us to discard these alternative mechanisms by isolating the oscillatory dynamics linked to the reactive dimension of inhibitory control and memory inhibition.

Conclusion

In this work, we explored the EEG oscillatory dynamics surrounding the control of unwanted memories during a “Think/No-Think” task. Using multivariate decoding of oscillatory dynamics, we showed that counter-intentional and interfering intrusive memories during attempts to prevent memories from entering awareness can be detected from the pattern of oscillatory responses using a perceptual model of intrusive visual images mimicking the counter-intentional reorientation of attentional focus. Our results suggest that memory and perceptual captures are subserved by a similar reflexive attentional mechanism reorienting the focus of attention toward counter-intentional input entering awareness. We also report that the Theta frequency band is a robust marker of both memory and attentional interference. Critically, we observed that the increase in Theta activity in response to the emergence of an intrusive memory, predicts future forgetting of the suppressed memory traces. This finding suggests that Theta increase is not only triggered by interfering memory activity but also reflects inhibitory control mechanisms that reactively purge memories out of awareness and inhibit the reactivation of associated engrams.

The uncontrolled reactivation of distressing memories is an initiating and aggravating symptom observable in many psychiatric disorders. By clarifying how reactive control contributes to memory suppression and inhibition, the present findings may offer insights into the mech-

anisms underlying intrusive symptoms in psychiatric disorders. Diminished suppression ability, specifically in reaction to intrusive and counter-intentional activity, should render people vulnerable to recurring and persisting intrusions that cascade into further distress. Further studies, however, would be needed to characterize the balanced contribution of reactive and proactive inhibition to such adverse outcomes.

Method

Participants

In total 27 participants (13 women) between the age of 18 and 35 years old were paid to participate. For the attention-to-attention decoding analysis, all 27 participants were included and analyzed. For the attention-to-TNT analysis, 6 participants had less than 4 intrusions that we could accurately decode during TNT. These participants were not included group-level analysis related to the TNT task. The study was approved by the local Ethics Committee (CPP Nord Ouest I, N° ID RCB: 2015-A01727), and all participants gave written consent. Participants were asked not to consume psychostimulants, drugs, alcohol or coffee during the hours preceding the experimental period and to avoid intense physical activity the day before.

Material

Attention. We selected 48 objects from the Bank of Standardized Stimuli (BOSS) (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014) and 32 scenes from the Nencki Affective Picture System (NAPS) database (Marchewka, Łukasz Żurawski, Jednoróg, & Grabowska, 2013; Riegel et al., 2015; Wierzba et al., 2015). For the Intrusion and Exploration condition, the stimuli were paired so each object was associated with the same condition during the 8 experimental blocks. The stimuli were paired and split into three subsets containing an equal number of Neutral and Disgust stimuli (12 in each group). Then, these three subsets were assigned to the Exploration, Intrusion and Non-intrusion condition in counterbalanced order between the participants.

TNT. We selected 72 objects from the Bank of Standardized Stimuli (BOSS) (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014) and 72 scenes

from the Nencki Affective Picture System (NAPS) database (Marchewka, Łukasz Żurawski, Jednoróg, & Grabowska, 2013; Riegel et al., 2015; Wierzba et al., 2015). All the items were different from the attentional procedure. The stimuli were paired and split into three subsets containing an equal number of Neutral and Disgust stimuli (12 in each group). Then, these three subsets were assigned to the Think, No-Think and Baseline condition in counterbalanced order between the participants.

Procedure

Step 1 - Learning and test/feedback of TNT pairs. Participants learned the 72 object-scene pairs used in the “Think/No-think” task plus the six training pairs. Participants first learned all object-scene pairs through a test-feedback cycle procedure. After studying all pairs for 6 s each, participants were given test trials presenting the object cue for a given pair for up to 4 s and asked whether they could recall and fully visualize the paired scene. If so, three scenes then appeared (one correct and two foils that were randomly taken from other pairs of the same emotional valence), and they received up to 4 s to select which scene went with the object cue. After selecting a scene or if the response window expired, a screen appeared for 1 s indicating whether the recognition judgment was correct or incorrect. In all cases (even if participants indicated that they could fully visualize the associated scene in the first step), each trial ended with the correct pairing appearing onscreen for 2.5 s. Participants were asked to use this feedback to increase their knowledge of the pair. Once all pairs had been tested, the percentage of total recall was displayed on the screen. If this percentage was inferior to 95% in at least one emotional condition, all the pairs were presented again in a randomized order. This procedure was repeated until the score was superior to 95%, thus ensuring correct encoding of the pairs and a comparable exposure to all the items.

Step 2 - Attentional task. The attentional task was designed to induce interference during a “Living/Non-living” classification task (see Fig. 1). Because this task was intended to model and simulate memory intrusion during the “Think/No-Think” task, we used a similar presentation paradigm. First, an image appeared in the center of the screen, surrounded by a red or a green box. For red trials, participants were instructed to determine whether an object (Non-living) or an animal (Living) was presented and to provide answers as fast and as accurately as possible. Half of the time, an interfering scene appeared in the background 200 ms after the onset of the object presentation. For green trials, an image was systematically displayed in the background 200 ms after the first stimulus presentation participants. Participants were instructed to determine whether humans (Living) were represented or not

(Non-living) and to provide answers as fast and as accurately as possible.

The attentional task was performed before the “Think/No-Think” task inside the Faraday’s cage. Participants were seated comfortably in a dimly lit room during the whole experiment at a distance from 90 cm to 100 cm from the screen and were instructed to minimize blinking and moving during recording. Stimulus presentation was controlled by Eprime Pro on a 17”LCD screen with a 1280 x 1024 resolution. To avoid exploratory eye movements, each picture was displayed inside a 400 x 400 pixels square. For each trial, the central picture depicting the object with a red or a green boundary was first presented on a black background. The square contour was colored in green or in red. Green boxes were associated with the Exploration condition and Red boxes with Intrusion and Non-Intrusion (no difference until the presentation of the intrusive picture on the screen). For Intrusion and Exploration trials, another larger picture appeared in the background 200 ms after the cued-object presentation. Intrusive and exploratory pictures were presented in the background with a dimension of 1024 x 768 pixels. Participant had to give a feedback concerning the “Living/No-living” categorization (for Exploration trials the target was the background picture, for Intrusion and Non-intrusion the target was the central object). Two responses were possible: **1**-Living, **2**-Non-living. All the images disappeared when the participant provided the response. This task was divided into 8 experimental blocks and lasted 27.05 ± 3.76 minutes (mean \pm standard deviation). Each block consisted in the presentations of 16 Exploration, 16 Intrusion and 8 Non-intrusion trials presented in random order. After two consecutive blocks, a message was displayed in a black screen indicating to the participant that he/she could rest. The participant decided when he/she wanted to resume the experiment by pressing any button on the response box. Note that participants underwent a training procedure in order to be familiarized with the attentional task before they enter into the Faraday’s cage. For this training, they performed 12 trials (4 Intrusion, 4 Non-intrusion and 4 Exploration conditions). The 12 objects and the 8 scenes we used were completely new for the participant were not part of the Think/No-Think material or of the Attentional procedure material. This familiarization was performed in front of a computer and the participant had to provide answers through the numerical keyboard (“1”: Living; “2”: Non-living). This order was consistent with the disposition of the response button inside the Faraday’s cage.

Step 3 - TNT criterion test. At this point, participants viewed a brief reminder of all of the studied pairs (3 s each), during which they were asked once again to reinforce their

knowledge of the pairings. Following this pair refresher, a final criterion test was proposed where participants had to recall the correct image in a similar way to the test/feedback procedure, but this time without feedback and only once. This allowed to isolate items forgotten before the TNT task and to exclude them from the analysis.

Step 4 - Think/no-think task. All the subjects performed the task inside a Faraday's cage after the attentional task and criterion test. The TNT object cues were displayed inside a 400 x 400 pixels square. Green and red square denoted the "Think" and "No-Think" condition respectively. The TNT task was divided into 8 TNT blocks, each 10–11 min in length. Each block consisted in the presentations of the 24 "Think" and 24 "No-think" items (12 Neutral, 12 Disgust each), yielding, across the 8 blocks, 256 trials for Neutral and 256 for Disgust conditions. Note that, although our design included both Neutral and Emotional material, here we could not separate these condition to avoid any drop in the number of total intrusions which would have prevented the decoding analysis. Cues appeared for 3000 ms either framed in green or red, centered on a grey background. On "Think" trials, the cue was bounded by a green box, and participants were told to generate an image of the associated scene as detailed and complete as possible. On "No-think" trials, the cue was bounded by a red box, and participants were told that it was imperative to prevent the scene from coming to mind at all and that they should fixate and concentrate on the object-cue without looking away. During red-cued trials, participants were asked to block thoughts of the scene by blanking their mind and not by replacing the scene 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 offset of each of the Think or No-Think trial cues, participants reported if the associated scene had entered awareness by pressing one of the response button. Three responses were possible: **1**-No intrusion at all, if the image did not intruded, **2**-The image intruded only for a short amount of time and/or was successfully controlled, **3**-The image intruded vividly for a large amount of time. Participants had up to 2 seconds to make this rating, so they were instructed and trained to make this quickly without thinking about the associated picture. These "intrusion ratings" were used to isolate trials with intrusive memories and quantify their occurrence. Specifically, we used participants' responses to classify each trial as either having an intrusion (i.e. a "briefly" or "frequent" response) or not (a "never" response) in a binary fashion. This phase lasted 41.77 ± 3.80 minutes (mean \pm standard deviation). After two consecutive blocks, a message was displayed in a black screen indicating to the participant that he/she could rest. The participant decided when he/she wanted

to resume the experiment by pressing any button on the response box. After the first 2 blocks, the experimenter entered in the Faraday's cage with the participant to propose a last questionnaire concerning the performances on the TNT task and ensure that the participant is performing it correctly. The same "diagnostic" questionnaire was also used during the training and instruction to TNT task, which was given right after the learning and test-feedback cycles (step 1) and before the participants enter the Faraday's cage.

Step 5 - Recall of TNT pairs. After this procedure, aftereffects of memory suppression were examined via a cued-recall task on all the object-scene pairs, including the 24 Baseline scenes omitted from the think/no-think task. Because these items were trained at the same time as the Think and No-Think items but did not participate in the "Think/No-think" task, they provide a baseline estimate of memorability of the scenes, given that they are similarly old but neither suppression nor retrieval had been performed on them. During trials of this cued-recall task, the cue object was presented in the center of the screen for 5 seconds and participant were told to visually recall the associated picture. If they could recall the associated scene, they were told to press the button under their right index finger and then had up to 15 sec to verbally describe the scene in as much detail as possible. The descriptions were recorded. If they could not recall the associated scene, they had to press the button under their right middle finger, and the next object was presented to them. Participants' descriptions were checked to ensure that they corresponded to the real pictures and could not be confounded with any other. On average, participants made false recall for 1% of the images.

EEG

EEG recording. EEG activity was recorded continuously by GES 300 amplifier (Electrical Geodesics, INC.) using an EGI Hydrocel Geodesic Sensor Net (HGSN-128) with dense array of 128 Ag/AgCl sensors (Tucker 1993). Impedances were kept under 100 k Ω (Ferree 2001), and EEG channel was referenced to a vertex reference Cz and ground to CPPZ (fixed by the EGI system). The signal was sampled at 20 kHz frequency with a 24-bit A/D and was online (hardware) amplified and low pass filtered at 4 KHz. The signal was filtered by a Butterworth low pass Finite Inpluse Response (FIR) filter at 500 Hz and sub-sampled at 1 KHz. Electro-oculogram was recorded using four electrodes placed vertically and horizontally around the eyes. EEG data were processed offline using Netstation 4.4.2 (Electrical Geodesics Inc.,

Eugene, OR, USA). The signal was filtered using a 1 Hz Kaiser FIR first order high-pass filter (which ensures a linear phase and no distortion in the bandwidth) in order to discard DC and very slow waves.

Preprocessing. EEG analyses were performed using the version 0.16.1 of the MNE library (Gramfort, 2013; Gramfort et al., 2014) implemented in Python 3.6 and followed recent recommended analysis steps (Andersen, 2018; Jas et al., 2018). We removed 13 peripheral electrodes (F11, FT11, T9, TP11, P09, I1, I2, P010, TP12, T10, FT12, F12), out of the 115 EEG electrodes, due to global bad impedance quality or no contact with the scalp at all. We low-pass filtered the raw data at 40 Hz cutoff frequency using a Finite Impulse Response (FIR) filter and referenced to a common average. On the remaining 102 channels, sensors showing no signal, constant white noise or intermittent signal, as observed during the EEG recording, were interpolated. Cue onset was adjusted for the delay (+15ms) between trigger generation and image appearance on the screen in the Faraday’s cage, measured using photodiodes during preliminary tests. We then segmented the trials from -1500ms to 2500ms for the attention task and from -1500ms to 4000ms around the cue onset for the TNT task. We detected and corrected artifacts using the Autoreject library (Jas, Engemann, Bekhti, Raimondo, & Gramfort, 2017) with local thresholds detection and the Bayesian optimization method. Finally, we applied ICA correction to reduce the remaining eye blink artifacts based on one manually selected EOG channel.

Time-frequency analysis. We extracted time-frequency representations for attention and TNT task using 27 Morlet wavelets from 3 to 30 HZ. The number of cycles was defined as half of the extracted frequency. After this procedure, we cropped the epochs from -0.5 s to 1.5s for the attention task and from -0.5 s to 3.0 s for the TNT task to exclude edge artifacts. For the decoding procedure, we corrected for baseline by computing the percentage of change as compared to the interval ranging from -500 ms to 0 ms (percent method). For the time-frequency analysis time-locked to memory intrusion, we calculated the logarithm of the power relative to the power during the baseline (-500 ms to 0 ms). In both cases, the resulting time-frequency power was then down-sampled to 20 ms time-bins before further analysis.

Multivariate decoding. We used random forest classifiers (Breiman, 2001) as implemented

in scikit-learn (Pedregosa et al., 2011) to decode the conditions during EEG recordings. Random forest algorithm is an ensemble method that combines many decision trees (Breiman, 1993) into a single model. Each of the trees is built on bootstrap samples of the training data, and each split within a particular tree is made upon a random subset of the predictors. Advantages are that random forest can handle very high dimensional dataset, even when containing a small number of observations or highly correlated predictor variables and are less likely to overfit. Because trials were excluded based on the autoreject artifact detection and trimmed for aberrant reaction times (> 3000 ms and < 400 ms), the attentional dataset contained different numbers of trials for each class. Although our design incorporated both Neutral and Disgusting intrusive images both in the attentional and TNT procedure, here we combined these stimuli altogether in order to improve the performance of the multivariate decoders. During the TNT procedure, the number of intrusions varied between participants and was on average lower than. Non-intrusion condition. We, therefore, specified an unbalanced *classweight* parameter for our analysis and used 500 estimators. Data were normalized through the *StandardScaler()* function. All the other parameters were set on default values.

First, we trained decoders to classify Intrusive and Non-intrusive trials from the attentional task. This was realized through the *SlidingEstimator()* function in MNE Python. The predictors at each time points (-500ms to 1000ms from the central image presentation) were the Theta power from the 102 selected EEG electrodes. In order to increase signal-to-noise ratio as well as to reduce computation time, the signal was down-sampled to 20 ms time bins. To control for over-fitting and ensure that the classifiers were never tested on data it was trained on, the classification was evaluated through a 10 fold cross-validation framework: each classifier was trained on 9/10 of the data and tested on 1/10 of the remaining data-set. The performances of the 10 predictors, indexed by the area under the ROC curve, were then averaged and compared to the chance level (0.5) for each time point using a permutation-based cluster one sample t-test. This allows to control for multiple testing and take into account the time dependence of the data.

To decode the memory intrusion during the “Think/No-Think”, we relied on the temporal generalization method (King & Dehaene, 2014; Dehaene & King, 2016) to infer the occurrence of similar Intrusive patterns during this procedure as learned from the attentional task. We analyzed here only No-Think trials, that we considered as intrusive if the participant either reported a moderate or intense intrusion (response **2** or response **3**, see the

TNT procedure) We trained 5 decoders in the 5 down-sampled time-points from 0 to 100 ms following the intrusive image presentation during the attention procedure. This interval was selected based on our observations of the time-resolved decoding performances from the previous procedure. Then, each decoder was used to compute the probability associated with the Intrusion condition for every time point of the No-Think trials (from -500 ms to 3000 ms from the reminder presentation). This measure ranges from 0 (low confidence for the Intrusion condition) to 1 (high confidence for the Intrusion condition). We then used this prediction probability to infer the occurrence of an intrusion. To improve signal-to-noise ratio, the resulting probabilities were smoothed using a Gaussian filter implemented in the *gaussian_filter1d()* function (sigma=1) from the SciPy Python package (Jones, Oliphant, Peterson, & others, 2001–). We then looked for peaks in the probabilities time course as an indicator of a possible intrusion. Peaks were extracted using the *indexes()* function from the PeakUtils Python package (Lucas Hermann Negri and Christophe Vestri, 2017). Each prediction probability higher than a threshold t was interpreted as an intrusion and the trial was labeled as an intrusive one. Only trials with 0 peaks were labeled as Non-intrusive. A minimum of 800 ms between consecutive peaks was specified in the function. We systematically looked for the best threshold by testing each value between 0.5 and 1 with steps of 0.02. All the predicted labels were then compared to the subjective report provided by the participant and the decoding accuracy was assessed through the area under the ROC curve. For each participant, we tested the decoding performances of the 5 time-bins for every threshold combinations and conserved the optimum combination.

Statistics

Statistical analyses were realized using JASP (v 0.9 <https://jasp-stats.org/>), the SciPy (Jones, Oliphant, Peterson, & others, 2001–) and the Pingouin Python packages (<https://raphaelvallat.github.io/pingouin/>). EEG data were analyzed using the MNE library (Gramfort, 2013). We also used the PeakUtils Python package (Lucas Hermann Negri and Christophe Vestri, 2017) to detect peaks decoding probabilities. Graphics were realized with MNE dedicated functions and the Seaborn python package (<https://github.com/mwaskom/seaborn>). We controlled for possible false positive due to a large number of comparisons in time-, frequency and spatial plots using nonparametric statistical testing (Maris & Oostenveld, 2007). We selected clusters of time, time-frequency or electrode that consecutively passed a specified threshold (threshold p -value=0.005) in the 1 or 2 dimensions of interest. The t -values of

each sample inside the resulting clusters were then summed and compared to the maximum cluster statistics obtained after 1024 random permutations. The comparison between this t-value and the random permutation distribution provided a p -value that is reported and interpreted as the cluster p -value.

The scripts used in this article will be made available on the following link: https://github.com/LegrandNico/intrusion_decoding/.

Supplementary materials

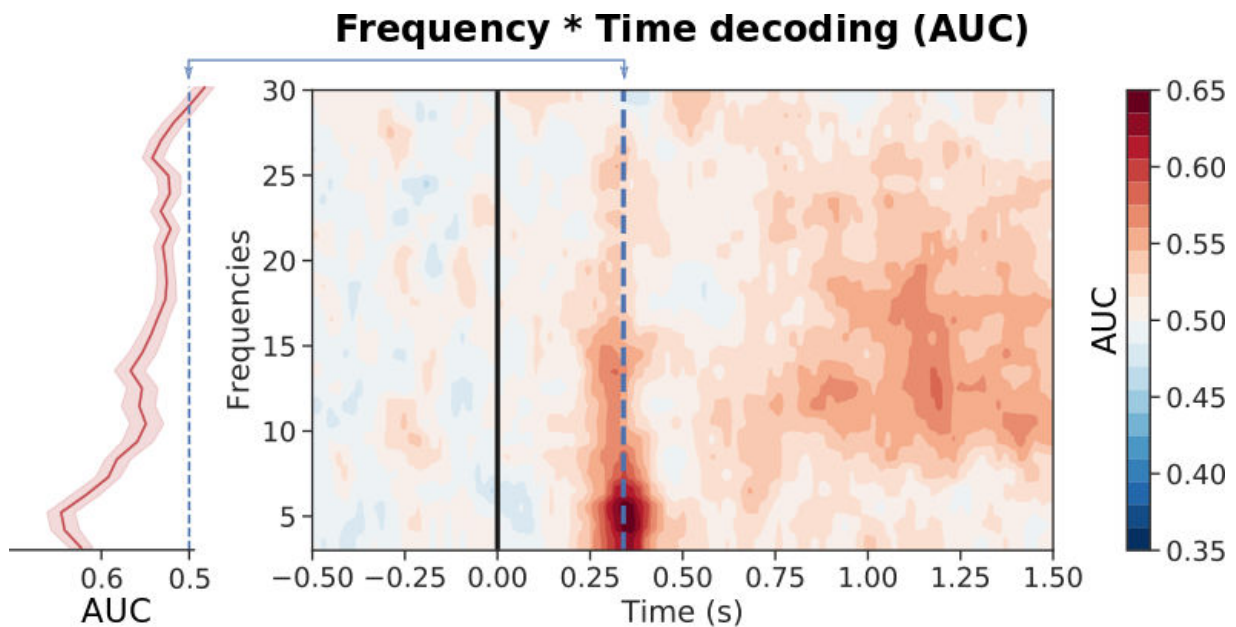


Figure 4: Decoding perceptual intrusions across time and frequency

Supplementary material 1, see Figure 4: *To confirm that the interference during the classification task was preferentially reflected in Theta (3-8 Hz) activity, we applied our decoding approach across all frequency bands between 3 and 30 Hz. The highest decoding score was achieved when the model was trained on the 5-6 Hz frequency band 330 ms after the onset of the initial object presentation.*

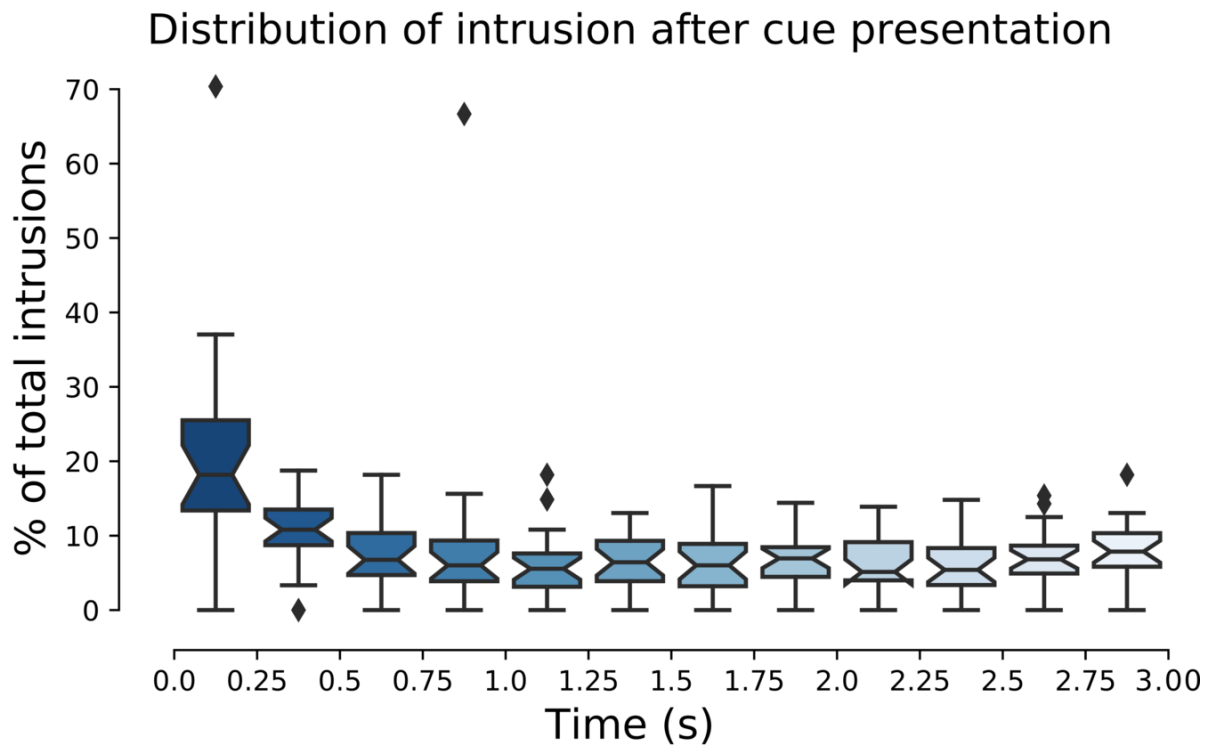


Figure 5: Distribution of intrusions for No-Think trials.

Supplementary material 2, see Figure 5: For each participant, we extracted the time of decoding for No-Think trials with explicit intrusion report. Then, we computed the proportion of intrusions inside bins of 0.25 s relative to the overall intrusion detection. Results show that the first 0.5 s contain on average more detected intrusions than the rest of the trials. However, the data also suggest that around 80% of total intrusion was detected after the first 0.25 s.

References

- 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. <http://doi.org/10.1016/j.actpsy.2007.12.004>
- Anderson, M. C., & Huddleston, E. (2011). Towards a Cognitive and Neurobiological Model of Motivated Forgetting. In *True and False Recovered Memories* (pp. 53–120). Springer New York. http://doi.org/10.1007/978-1-4614-1195-6_3
- Anderson, M. C., Bunce, J. G., & Barbas, H. (2016). Prefrontal–hippocampal pathways underlying inhibitory control over memory. *Neurobiology of Learning and Memory*, *134*, 145–161. <http://doi.org/10.1016/j.nlm.2015.11.008>
- Anderson, M. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. <http://doi.org/10.1016/j.jml.2003.08.006>
- Hu, X., Bergström, Z. M., Gagnepain, P., & Anderson, M. C. (2017). Suppressing Unwanted Memories Reduces Their Unintended Influences. *Current Directions in Psychological Science*, *26*(2), 197–206. <http://doi.org/10.1177/0963721417689881>
- Anderson, M. C. (2004). Neural Systems Underlying the Suppression of Unwanted Memories. *Science*, *303*(5655), 232–235. <http://doi.org/10.1126/science.1089504>
- Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process. *Science*, *317*(5835), 215–219. <http://doi.org/10.1126/science.1139560>
- Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences*, *111*(13), E1310–E1319. <http://doi.org/10.1073/pnas.1311468111>
- Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories. *The Journal of Neuroscience*, *37*(27), 6423–6441. <http://doi.org/10.1523/jneurosci.2732-16.2017>
- Benoit, R. G., & Anderson, M. C. (2012). Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories. *Neuron*, *76*(2), 450–460. <http://doi.org/10.1016/j.>

neuron.2012.07.025

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive Top-Down Suppression of Hippocampal Activity and the Purging of Intrusive Memories from Consciousness. *Journal of Cognitive Neuroscience*, *27*(1), 96–111. http://doi.org/10.1162/jocn_a_00696

Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to Forget. *Psychological Science*, *26*(5), 604–616. <http://doi.org/10.1177/0956797615569889>

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366–369. <http://doi.org/10.1038/35066572>

Bergström, Z. M., Anderson, M. C., Buda, M., Simons, J. S., & Richardson-Klavehn, A. (2013). Intentional retrieval suppression can conceal guilty knowledge in ERP memory detection tests. *Biological Psychology*, *94*(1), 1–11. <http://doi.org/10.1016/j.biopsycho.2013.04.012>

Bergström, Z. M., Velmans, M., de Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. *Brain Research*, *1151*, 119–133. <http://doi.org/10.1016/j.brainres.2007.03.014>

Chen, C., Liu, C., Huang, R., Cheng, D., Wu, H., Xu, P., ... Luo, Y.-J. (2012). Suppression of aversive memories associates with changes in early and late stages of neurocognitive processing. *Neuropsychologia*, *50*(12), 2839–2848. <http://doi.org/10.1016/j.neuropsychologia.2012.08.004>

Hanslmayr, S., Leipold, P., Pastotter, B., & Bauml, K.-H. (2009). Anticipatory Signatures of Voluntary Memory Suppression. *Journal of Neuroscience*, *29*(9), 2742–2747. <http://doi.org/10.1523/jneurosci.4703-08.2009>

Hellerstedt, R., Johansson, M., & Anderson, M. C. (2016). Tracking the intrusion of unwanted memories into awareness with event-related potentials. *Neuropsychologia*, *89*, 510–523. <http://doi.org/10.1016/j.neuropsychologia.2016.07.008>

Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain Research*, *1255*, 132–147. <http://doi.org/10.1016/j.brainres.2008.11.073>

- Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*(4), 726–737. <http://doi.org/10.1016/j.neuroimage.2009.06.051>
- Bergström, Z. M., de Fockert, J., & Richardson-Klavehn, A. (2009). Event-related Potential Evidence that Automatic Recollection Can Be Voluntarily Avoided. *Journal of Cognitive Neuroscience*, *21*(7), 1280–1301. <http://doi.org/10.1162/jocn.2009.21075>
- Ketz, N., O'Reilly, R. C., & Curran, T. (2014). Classification aided analysis of oscillatory signatures in controlled retrieval.. *Neuroimage*, *85 Pt 2*, 749–60.
- Waldhauser, G. T., Bäuml, K.-H. T., & Hanslmayr, S. (2014). Brain Oscillations Mediate Successful Suppression of Unwanted Memories. *Cerebral Cortex*, *25*(11), 4180–4190. <http://doi.org/10.1093/cercor/bhu138>
- Depue, B. E., Ketz, N., Mollison, M. V., Nyhus, E., Banich, M. T., & Curran, T. (2013). ERPs and Neural Oscillations during Volitional Suppression of Memory Retrieval. *Journal of Cognitive Neuroscience*, *25*(10), 1624–1633. http://doi.org/10.1162/jocn_a_00418
- Waldhauser, G. T., Dahl, M. J., Ruf-Leuschner, M., Müller-Bamouh, V., Schauer, M., Axmacher, N., ... Hanslmayr, S. (2018). The neural dynamics of deficient memory control in heavily traumatized refugees. *Scientific Reports*, *8*(1). <http://doi.org/10.1038/s41598-018-31400-x>
- Legrand, N., Etard, O., Vandavelde, A., Pierre, M., Viader, F., Clochon, P., ... Gagnepain, P. (2018). Does the heart forget? Modulation of cardiac activity induced by inhibitory control over emotional memories. <http://doi.org/10.1101/376954>
- Levy, B. J., & Anderson, M. C. (2012). Purging of Memories from Conscious Awareness Tracked in the Human Brain. *Journal of Neuroscience*, *32*(47), 16785–16794. <http://doi.org/10.1523/jneurosci.2640-12.2012>
- Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613–625. <http://doi.org/10.1038/nrn2459>
- Cabeza, R., Mazuz, Y. S., Stokes, J., Kragel, J. E., Woldorff, M. G., Ciaramelli, E., ... Moscovitch, M. (2011). Overlapping Parietal Activity in Memory and Perception: Evidence for the Attention to Memory Model. *Journal of Cognitive Neuroscience*, *23*(11), 3209–3217. http://doi.org/10.1162/jocn_a_00065

- Ciaramelli, E., Grady, C., Levine, B., Ween, J., & Moscovitch, M. (2010). Top-Down and Bottom-Up Attention to Memory Are Dissociated in Posterior Parietal Cortex: Neuroimaging and Neuropsychological Evidence. *Journal of Neuroscience*, *30*(14), 4943–4956. <http://doi.org/10.1523/jneurosci.1209-09.2010>
- Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828–1851. <http://doi.org/10.1016/j.neuropsychologia.2008.03.022>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. <http://doi.org/10.1016/j.tics.2014.04.012>
- Helfrich, R. F., & Knight, R. T. (2016). Oscillatory Dynamics of Prefrontal Cognitive Control. *Trends in Cognitive Sciences*, *20*(12), 916–930. <http://doi.org/10.1016/j.tics.2016.09.007>
- Klimesch, W. (2012). Alpha-band oscillations attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. <http://doi.org/10.1016/j.tics.2012.10.007>
- Clouter, A., Shapiro, K. L., & Hanslmayr, S. (2017). Theta Phase Synchronization Is the Glue that Binds Human Associative Memory. *Current Biology*, *27*(20), 3143–3148.e6. <http://doi.org/10.1016/j.cub.2017.09.001>
- Zhang, H., Watrous, A. J., Patel, A., & Jacobs, J. (2018). Theta and Alpha Oscillations Are Traveling Waves in the Human Neocortex. *Neuron*, *98*(6), 1269–1281.e4. <http://doi.org/10.1016/j.neuron.2018.05.019>
- Hanslmayr, S., Pastötter, B., Bäuml, K.-H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The Electrophysiological Dynamics of Interference during the Stroop Task. *Journal of Cognitive Neuroscience*, *20*(2), 215–225. <http://doi.org/10.1162/jocn.2008.20020>
- Womelsdorf, T., Johnston, K., Vinck, M., & Everling, S. (2010). Theta-activity in anterior cingulate cortex predicts task rules and their adjustments following errors. *Proceedings of the National Academy of Sciences*, *107*(11), 5248–5253. <http://doi.org/10.1073/pnas.0906194107>
- Rajan, A., Siegel, S. N., Liu, Y., Bengson, J., Mangun, G. R., & Ding, M. (2018). Theta

Oscillations Index Frontal Decision-Making and Mediate Reciprocal Frontal–Parietal Interactions in Willed Attention. *Cerebral Cortex*. <http://doi.org/10.1093/cercor/bhy149>

Kam, J. W. Y., Solbakk, A.-K., Endestad, T., Meling, T. R., & Knight, R. T. (2018). Lateral prefrontal cortex lesion impairs regulation of internally and externally directed attention. *NeuroImage*, *175*, 91–99. <http://doi.org/10.1016/j.neuroimage.2018.03.063>

Popov, T., Popova, P., Harkotte, M., Awiszus, B., Rockstroh, B., & Miller, G. A. (2018). Cross-frequency interactions between frontal theta and posterior alpha control mechanisms foster working memory. *NeuroImage*, *181*, 728–733. <http://doi.org/10.1016/j.neuroimage.2018.07.067>

Oehr, C. R., Fell, J., Baumann, C., Rosburg, T., Ludowig, E., Kessler, H., ... Axmacher, N. (2018). Direct Electrophysiological Evidence for Prefrontal Control of Hippocampal Processing during Voluntary Forgetting. *Current Biology*, *28*(18), 3016–3022.e4. <http://doi.org/10.1016/j.cub.2018.07.042>

Mückschel, M., Dippel, G., & Beste, C. (2017). Distinguishing stimulus and response codes in theta oscillations in prefrontal areas during inhibitory control of automated responses. *Human Brain Mapping*, *38*(11), 5681–5690. <http://doi.org/10.1002/hbm.23757>

Bramson, B., Jensen, O., Toni, I., & Roelofs, K. (2018). Cortical Oscillatory Mechanisms Supporting the Control of Human Social–Emotional Actions. *The Journal of Neuroscience*, *38*(25), 5739–5749. <http://doi.org/10.1523/jneurosci.3382-17.2018>

Ferreira, C. S., Maraver, M. J., Hanslmayr, S., & Bajo, M. T. (2018). Theta oscillations show impaired interference detection in the elderly during selective memory retrieval. <http://doi.org/10.1101/388595>

Staudigl, T., Hanslmayr, S., & Bauml, K.-H. T. (2010). Theta Oscillations Reflect the Dynamics of Interference in Episodic Memory Retrieval. *Journal of Neuroscience*, *30*(34), 11356–11362. <http://doi.org/10.1523/jneurosci.0637-10.2010>

Töllner, T., Wang, Y., Makeig, S., Müller, H. J., Jung, T.-P., & Gramann, K. (2017). Two Independent Frontal Midline Theta Oscillations during Conflict Detection and Adaptation in a Simon-Type Manual Reaching Task. *The Journal of Neuroscience*, *37*(9), 2504–2515. <http://doi.org/10.1523/jneurosci.1752-16.2017>

Hanslmayr, S., Volberg, G., Wimber, M., Oehler, N., Staudigl, T., Hartmann, T., ... Bauml, K.-H. T. (2012). Prefrontally Driven Downregulation of Neural Synchrony Mediates Goal-

- Directed Forgetting. *Journal of Neuroscience*, *32*(42), 14742–14751. <http://doi.org/10.1523/jneurosci.1777-12.2012>
- Hanslmayr, S., Staudigl, T., Aslan, A., & Bäuml, K.-H. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive Affective, & Behavioral Neuroscience*, *10*(3), 329–338. <http://doi.org/10.3758/cabn.10.3.329>
- Wimber, M., Maaß, A., Staudigl, T., Richardson-Klavehn, A., & Hanslmayr, S. (2012). Rapid Memory Reactivation Revealed by Oscillatory Entrainment. *Current Biology*, *22*(16), 1482–1486. <http://doi.org/10.1016/j.cub.2012.05.054>
- Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in Neurosciences*, *37*(9), 480–490. <http://doi.org/10.1016/j.tins.2014.06.004>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. <http://doi.org/10.1016/j.jneumeth.2007.03.024>
- Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. <http://doi.org/10.3389/fnins.2013.00267>
- King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. <http://doi.org/10.1016/j.tics.2014.01.002>
- Dehaene, S., & King, J.-R. (2016). Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method. In *Research and Perspectives in Neurosciences* (pp. 85–97). Springer International Publishing. http://doi.org/10.1007/978-3-319-28802-4_7
- Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bäuml, K.-H. T. (2011). The Relationship between Brain Oscillations and BOLD Signal during Memory Formation: A Combined EEG-fMRI Study. *Journal of Neuroscience*, *31*(44), 15674–15680. <http://doi.org/10.1523/jneurosci.3140-11.2011>
- Chun, M. M., & Johnson, M. K. (2011). Memory: Enduring Traces of Perceptual and Reflective Attention. *Neuron*, *72*(4), 520–535. <http://doi.org/10.1016/j.neuron.2011.10.026>
- Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying Object Cat-

- egories from Event-Related EEG: Toward Decoding of Conceptual Representations. *PLoS ONE*, 5(12), e14465. <http://doi.org/10.1371/journal.pone.0014465>
- Robinson, A. K., Venkatesh, P., Boring, M. J., Tarr, M. J., Grover, P., & Behrmann, M. (2017). Very high density EEG elucidates spatiotemporal aspects of early visual processing. *Scientific Reports*, 7(1). <http://doi.org/10.1038/s41598-017-16377-3>
- Dijkstra, N., Mostert, P., de Lange, F. P., Bosch, S., & van Gerven, M. A. J. (2018). Differential temporal dynamics during visual imagery and perception. *ELife*, 7. <http://doi.org/10.7554/elife.33904>
- King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, 92(5), 1122–1134. <http://doi.org/10.1016/j.neuron.2016.10.051>
- Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2005). Relevance of EEG alpha and theta oscillations during task switching. *Experimental Brain Research*, 170(3), 295–301. <http://doi.org/10.1007/s00221-005-0211-y>
- Cooper, P. S., Wong, A. S. W., Fulham, W. R., Thienel, R., Mansfield, E., Michie, P. T., & Karayanidis, F. (2015). Theta frontoparietal connectivity associated with proactive and reactive cognitive control processes. *NeuroImage*, 108, 354–363. <http://doi.org/10.1016/j.neuroimage.2014.12.028>
- Kerrén, C., Linde-Domingo, J., Hanslmayr, S., & Wimber, M. (2018). An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval. *Current Biology*. <http://doi.org/10.1016/j.cub.2018.08.065>
- Schönauer, M., Alizadeh, S., Jamalabadi, H., Abraham, A., Pawlizki, A., & Gais, S. (2017). Decoding material-specific memory reprocessing during sleep in humans. *Nature Communications*, 8, 15404. <http://doi.org/10.1038/ncomms15404>
- Schreiner, T., Doeller, C. F., Jensen, O., Rasch, B., & Staudigl, T. (2018). Theta Phase-Coordinated Memory Reactivation Reoccurs in a Slow-Oscillatory Rhythm during NREM Sleep. *Cell Reports*, 25(2), 296–301. <http://doi.org/10.1016/j.celrep.2018.09.037>
- Eschmann, K. C. J., Bader, R., & Mecklinger, A. (2018). Topographical differences of frontal-midline theta activity reflect functional differences in cognitive control abilities. *Brain and Cognition*, 123, 57–64. <http://doi.org/10.1016/j.bandc.2018.02.002>

- Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220–238. <http://doi.org/10.1111/j.1469-8986.2011.01293.x>
- Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: New insights into the localization meaning and network organization. *NeuroImage*, *84*, 159–168. <http://doi.org/10.1016/j.neuroimage.2013.08.028>
- Verguts, T. (2017). Binding by Random Bursts: A Computational Model of Cognitive Control. *Journal of Cognitive Neuroscience*, *29*(6), 1103–1118. http://doi.org/10.1162/jocn_a_01117
- Mizuhara, H., & Yamaguchi, Y. (2007). Human cortical circuits for central executive function emerge by theta phase synchronization. *NeuroImage*, *36*(1), 232–244. <http://doi.org/10.1016/j.neuroimage.2007.02.026>
- Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in Cognitive Sciences*, *13*(7), 293–301. <http://doi.org/10.1016/j.tics.2009.04.005>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS) a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773. <http://doi.org/10.1371/journal.pone.0010773>
- Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, *9*(9), e106953. <http://doi.org/10.1371/journal.pone.0106953>
- Marchewka, A., Łukasz Żurawski, Jednoróg, K., & Grabowska, A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610. <http://doi.org/10.3758/s13428-013-0379-1>
- Riegel, M., Łukasz Żurawski, Wierzbą, M., Moslehi, A., Łukasz Klocek, Horvat, M., ... Marchewka, A. (2015). Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). *Behavior Research Methods*, *48*(2), 600–612. <http://doi.org/10.3758/s13428-015-0620-1>
- Wierzbą, M., Riegel, M., Pucz, A., Leśniewska, Z., Dragan, W. Ł., Gola, M., ... Marchewka,

- A. (2015). Erotic subset for the Nencki Affective Picture System (NAPS ERO): cross-sexual comparison study. *Frontiers in Psychology*, 6. <http://doi.org/10.3389/fpsyg.2015.01336>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86, 446–460. <http://doi.org/10.1016/j.neuroimage.2013.10.027>
- Andersen, L. M. (2018). Group Analysis in MNE-Python of Evoked Responses from a Tactile Stimulation Paradigm: A Pipeline for Reproducibility at Every Step of Processing Going from Individual Sensor Space Representations to an across-Group Source Space Representation. *Frontiers in Neuroscience*, 12. <http://doi.org/10.3389/fnins.2018.00006>
- Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations Quality Assessments, and Good Practices. *Frontiers in Neuroscience*, 12. <http://doi.org/10.3389/fnins.2018.00530>
- Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 159, 417–429. <http://doi.org/10.1016/j.neuroimage.2017.06.030>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <http://doi.org/10.1023/a:1010933404324>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Breiman, L. (1993). *Classification and regression trees*. New York: Chapman & Hall.
- Jones, E., Oliphant, T., Peterson, P., & others. (2001–2001–). SciPy: Open source scientific tools for Python. Retrieved from <http://www.scipy.org/>
- Lucas Hermann Negri and Christophe Vestri, month=S., year=2017. (2017). lucashn/peakutils: v1.1.0. <http://doi.org/10.5281/zenodo.887917>

4.1 Mémoires individuelles, mémoires collectives

Nous avons abordé dans les deux parties précédentes l'impact de l'inhibition du rappel des souvenirs sur la mémoire épisodique. Les études qui ont été présentées nous ont conduits à reconsidérer les mécanismes de l'oubli volontaire en les intégrant dans une dimension incarnée, sur le plan de la physiologie des émotions, et dynamique pour ce qui concerne l'inhibition des intrusions mnésiques. L'oubli est une composante importante de la mémoire qui a longtemps été prise pour son contraire ou son échec (Eustache, 2014). La mise à l'écart de cette fonction est encore plus marquée si l'on se place à une échelle encore plus large, celle de l'étude par les sciences humaines et sociales de la mémoire collective (Lavabre, 2013) qui va nous intéresser maintenant. L'oubli tient pourtant une place essentielle si nous cherchons à comprendre la dynamique de construction d'un système de mémoire sur le temps long. Nous savons l'Homme intégré au sein de sociétés et de cultures qui, par-delà la conservation du souvenir dans les systèmes de mémoire biologiques décrits par les neurosciences, peuvent aussi conserver la trace de temps plus long et influencer les représentations de générations successives. Nous savons que le cerveau, et les pratiques scientifiques qui en abordent certaines dimensions, doivent pleinement intégrer ces contextes sociaux aux variables de leurs études (Peschanski, 2013). Ces structures sont aussi celles des sciences humaines, pour qui l'hégémonie actuelle des neurosciences dans l'étude des comportements humains est alternativement une chance, l'aboutissement d'un processus et un défi pour l'affirmation de la singularité de leur objet. Dans la partie suivante, nous allons tenter de montrer que l'oubli, que nous avons décrit pour le moment à l'échelle de l'individu, est une autre étape dans la construction de représentations collectives plus larges. Et si ces représentations sont souvent un soutien pour les souvenirs, nous verrons qu'une forme d'oubli, ou tout du moins d'occultation, peut intervenir dans la mémoire collective et osciller en fonction de différents contextes historiques et sociaux.

4.1.1 Mémoire collective : des neurosciences à l'histoire

La mémoire a une fonction : conserver les représentations d'événements dans le temps pour en livrer une dimension compréhensible dans le présent. Dire que cette fonction s'applique aujourd'hui par analogie à des systèmes d'une exceptionnelle variété, tant dans le domaine du vivant, de l'intelligence artificielle ou dans les concepts même qu'utilisent les sciences humaine et sociales, ne ferait qu'effleurer la question. Mais c'est pourtant bien par un glissement métaphorique que le corps social s'est vu doter, avec la naissance de la sociologie, d'attributs psychiques, d'une *conscience* et aussi d'une *mémoire*. L'isomorphisme qui va du corps humain au corps collectif n'est pas récent. Dans *La République*, Platon procédait déjà à une forme de psychosociologie lorsqu'il décrivait d'un même mouvement les parties de l'âme humaine et

les classes de la cité idéale (Platon, -315/2016). C'est au moyen de l'analogie, en partant de représentations sensibles et intuitives qu'il était alors possible de penser les dynamiques structurant les groupes humains. Ces articulations sont inhérentes à l'épistémologie des sciences sociales. L'utilisation de l'analogie se retrouve aussi dans les problèmes épistémologiques qui se posent à l'interface de ces disciplines et des neurosciences. Le mouvement qui va du neurone au cerveau est le même, par bien des aspects, que celui qui part de l'individu pour aller au corps social : il s'agit à chaque fois de questionner la complexité supplémentaire qu'ajoute le tout sur la somme des parties. Mais ici, nous pouvons aussi voir que cette réutilisation des fonctions cognitive se redoublait d'une nouvelle qualification qu'il faut comprendre : la mémoire était devenue "*collective*", et c'est le sens de ce mot qu'il convient d'interroger dans un premier temps.

La mémoire collective se définit actuellement comme un ensemble de représentations du passé constitutives de l'identité d'un groupe (qu'il s'agisse de quelques personnes ou de communautés très larges), sélectionnées et partagées par l'ensemble de ses membres qui sont connectés par une histoire commune (Hirst et al., 2018). Le lien géographique est implicitement considéré comme un préalable de la constitution de telles communautés. C'était particulièrement le cas lors des premières définitions qu'en a proposé le sociologue français Maurice Halbwachs (Halbwachs, 1925) à une époque où l'idée de la nation et de l'État jouaient un rôle très important. Cette dimension a pourtant tendance à être de plus en plus écartée aujourd'hui alors que les communautés peuvent être plus diffuses et éclatées, notamment à travers l'usage des réseaux sociaux (Meshi et al., 2015). La mémoire collective est aujourd'hui un objet d'étude partagé par de nombreuses disciplines, incluant l'histoire, la sociologie, la psychologie sociale et cognitive et, de façon très récente, les neurosciences. On considère aussi que son étude peut se faire selon une approche **ascendante** (de l'individu aux cadres sociaux) ou **descendante**. Bien que les questions portées par ces domaines soient multiples, les mécanismes concernant, d'une part la **formation des souvenirs collectifs** à partir d'événements récents, et d'autre part la structure des **représentations d'événements distants** sont deux thèmes qui traversent invariablement ce champ d'étude.

La psychologie cognitive s'est emparée depuis quelques années de l'étude de cette dimension de la mémoire (Hirst et al., 2018; Roediger and Abel, 2015; Hirst and Manier, 2008; Barnier and Sutton, 2008) en intégrant à des degrés variables les formes de partages des souvenirs que nous avons évoquées. Il importe dans un premier temps de dissocier ce nouveau champ de recherche d'un autre domaine, extrêmement actif depuis une quinzaine d'années et relativement proche, qui s'inscrit dans le développement naturel des neurosciences et de la psychologie expérimentale, à savoir l'étude de la « *dimension sociale* » de la mémoire. Prise comme objet par les neurosciences et la neuropsychologie, la mémoire comporte déjà dans sa théorisation de nombreuses dimensions. Le versant social - qui a donné lieu au *tournant social* actuel (2014) est l'une des étapes d'un processus qui a considérablement ouvert l'économie de la mémoire à un nombre croissant d'influences, des émotions à la

théorie de l'esprit, et en intégrant de plus en plus les interactions interindividuelles dans un sens large. Ce changement est particulièrement marquant si on le contraste avec les travaux de certains pères de la psychologie qui avaient méthodiquement isolé l'individu de toute influence sociale pour accéder à une forme épurée des processus de mémorisation. L'exemple du psychologue allemand Hermann Ebbinghaus est à ce titre emblématique puisque celui-ci expérimentait la rétention de listes de syllabes dénuées de signification et d'émotions directement sur lui-même (Ebbinghaus, 1885/2010). Pourtant, cette inflexion que représente le tournant social s'est faite dans une certaine autonomie vis-à-vis des interactions possibles avec la sociologie ou l'histoire – porteuses initiales du concept de mémoire *collective* qui nous intéressera ici. Nous avons décrit plus en détail ce mouvement que représente l'intégration des neurosciences aux *memory studies* contemporaines dans une publication que nous plaçons en annexe de ce travail (Legrand et al., 2015). Dans sa compréhension actuelle, le concept de mémoire collective fait intrinsèquement intervenir différentes temporalités (celles des individus et des corps sociaux), différentes modalités de représentation (des neurones à l'inscription culturelle de souvenirs) et autant de disciplines qui peuvent tenter d'approcher certaines composantes de ces interactions. Ces disciplines qui ont abordé cette notion ont opéré dans des cadres épistémologiques et parfois même historiques distincts qui ont influencé la compréhension de ce phénomène. Nous proposons dans la partie suivante de distinguer trois notions de "mémoires collectives". Ce partage, qui peut lui-même être débattu, présente l'intérêt d'explicitier trois définitions du terme "*mémoire*", et trois compréhensions de ce mouvement ou de cette différence qui l'a faite devenir "*collective*".

4.1.2 Trois variantes des mémoires collectives

Le premier sens que l'on pourra donner à la mémoire collective, et sans doute celui qui aura la plus grande importance dans son utilisation par les neurosciences, désigne le **partage de représentations**. Les individus sont alors liés par le souvenir en commun, traces de l'événement vécu inscrites dans chaque individu sur le plan neural avec un niveau de similarité qui autorise à voir dans la diversité des représentations individuelles les exemplaires reproduits d'un même souvenir. De façon déterminante, les neurosciences ont réussi relativement récemment à s'emparer de cette thématique (Gagnepain et al., *forthcoming*). Ces avancées marquent un dialogue direct et clair entre les modèles proposés par l'étude en sciences humaines et sociales de la mémoire collective et la neurobiologie. Ce résultat est le fruit d'une avancée sur le plan théorique qui a rendu possible un tel dialogue transdisciplinaire. Toutefois, l'idée d'une influence culturelle ou sociale sur les structures neurales a également quelques antécédents dans l'histoire récente des sciences. Il faut donc aussi voir dans ces avancées le résultat d'un travail méthodologique, celui qui a permis de fournir les outils mathématiques permettant de quantifier de façon robuste cette notion de similarité entre les états mentaux de différents individus. Ceci a notamment été permis par le développement des Analyses de Similarité Représentationnelles (Kriegeskorte, 2008) que nous utiliserons dans la prochaine étude. Cette forme de *souvenirs en commun* ne change pas fondamentalement la fonction même

de la mémoire humaine puisque les mécanismes qui en assurent l'encodage et la récupération sont déjà connus des neurosciences. Pourtant, nous pouvons déjà distinguer ce qui relèverait de la représentation partagée (Chen et al., 2016; Edelson et al., 2011), le partage de patterns représentationnels suite à l'exposition d'un groupe d'individus à la même expérience, d'une mémoire collective qui ajoute un référentiel culturel, identitaire et théorique. Ainsi, toutes les représentations qui sont partagées par un groupe de personnes n'intègrent pas systématiquement la mémoire collective de ce groupe : elles doivent véhiculer une signification qui participe au maintien de la cohésion du groupe, à son histoire, et aussi à l'interprétation même de nouveaux événements.

La seconde acceptation désignera non plus les représentations partagées mais les représentations en partage. Ce sens reprend en partie la distinction au sein de la mémoire collective d'une **mémoire culturelle** et d'une **mémoire communicationnelle**, ce qui a été proposé par l'égyptologue allemand Jan Assman (Assmann and Czaplicka, 1995; Assmann, 2011). Au-delà de la mémoire individuelle telle que nous pouvons l'appréhender par l'étude du cerveau, une dimension communicationnelle de la mémoire collective désignera l'ensemble des connaissances partagées par un groupe d'individus pendant une période historique déterminée et participera à ancrer l'identité sociale des individus au sein du groupe. Ces représentations sont pourtant limitées au temps historique considéré et ne peuvent plus se communiquer sans les individus qui les portent. Une mémoire culturelle au contraire sera inscrite dans un héritage transmissible à l'échelle de plusieurs générations et participera à ancrer l'identité culturelle des individus qui y sont exposés. Le passage de certaines représentations de la mémoire communicationnelle à la mémoire culturelle est un élément central du processus mémoriel. Selon cette définition, le souvenir est un élément extérieur à l'individu, il est inscrit dans l'environnement culturel auquel il est exposé. Le terme collectif désigne alors non plus ce qui est inscrit dans tous les systèmes de mémoire, mais ce qui est rendu accessible à tous les systèmes de mémoire, sans pour autant que ces représentations soient activement exploitées et aient une prégnance forte dans la mémoire communicationnelle. Il faut aussi nuancer cette idée de l'extériorisation de la dimension culturelle de la mémoire en soulignant que tout souvenir qui est amené à figurer dans ces ressources n'intègre pas nécessairement la mémoire collective. On distinguera une dimension **publique** et une dimension **privée** de ces souvenirs. Un album de photographies familiales pourrait ainsi se situer alternativement dans l'une ou l'autre de ces catégories en fonction de son contexte et du référentiel collectif que l'on considère : l'album est partagé au sein de la famille mais est une archive privée au regard d'autres groupes. L'idée selon laquelle la mémoire collective serait une sorte d'opposable à la mémoire individuelle ne va en réalité pas de soi. La mémoire collective ne consiste pas uniquement en ce constat passif du partage de certains souvenirs au sein d'une communauté, mais d'un acte volontaire de mise en commun de ces traces en fonction des événements qui les entourent.

Enfin, le dernier sens se rattachera à la notion peut-être plus débattue de **processus émergent**. La psychologie et la philosophie (Wilson, 2005) ont pu aborder la question de la mémoire collective en s'interrogeant sur le système de mémoire que constitue la société dans son ensemble, et l'influence des cadres sociaux sur le système cognitif à travers la théorie de l'*extended mind* (Clark and Chalmers, 1998). Selon cette approche, l'esprit ne pourrait être restreint à l'ensemble des processus opérant dans les limites du corps mais interagit avec des objets, outils ou personnes dans son fonctionnement qui doivent être compris comme faisant partie intégrante de son économie. A l'échelle d'un groupe, les interactions interpersonnelles, fondées sur les dynamiques cognitives des individus, peuvent faire émerger des comportements collectifs qui ne se réduisent pas à leurs composantes individuelles (Vlasceanu et al., 2018). Le souvenir est alors encodé dans une structure qui dépasse en quelque sorte les individus, ce à quoi réfère le caractère collectif de cette mémoire. Tout comme les comportements d'une foule ne sont pas ceux de la somme de ses individus, il est possible de concevoir une forme de mémoire propre aux dynamiques de vastes groupes que l'on ne saurait entièrement comprendre en intégrant uniquement le partage des représentations ou les représentations en partage.

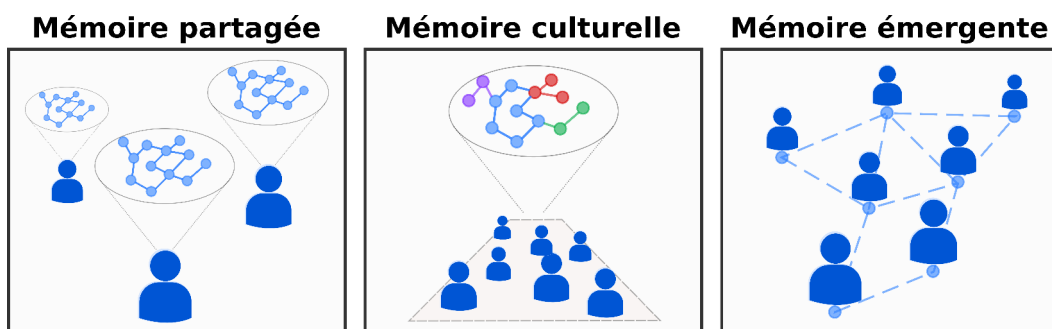


Fig. 4.1: Trois notions de mémoire collective. La mémoire collective peut renvoyer à trois définitions qui se distinguent par la nature des souvenirs qui sont considérés, leurs modalités d'accès et le sens de la dimension collective de la mémoire. La mémoire **partagée** désigne un ensemble de souvenirs inscrit dans l'individu sous forme de traces neurales qui entretiennent une relation de similarité avec les souvenirs d'autres membres du groupe qui sont liés par une expérience, une histoire ou une identité commune. La mémoire **culturelle** renverra à un ensemble de représentations inscrites dans l'environnement direct des individus et se distingue d'une mémoire communicationnelle utilisée activement dans le contexte historique considéré. Les souvenirs survivent aux individus et dispose de nombreux moyens d'accès différents, par exemple à travers différents médias. Les individus sont liés entre eux par l'environnement culturel qui les influence. La mémoire **émergente** est une propriété de l'ensemble formé par la connexion des individus entre eux. Les souvenirs correspondent à la récupération d'un état passé sous l'influence d'un événement.

Les trois définitions que nous venons d'évoquer montrent que la mémoire peut se faire "collective" de plusieurs façons. Si la mémoire elle-même ne semble pas démissionner de sa fonction de conservation de traces d'événements passés, ces traces sont conservées selon des modalités différentes. La psychologie et les neurosciences ont

évidemment un point d'accroche majeur dans la première définition que nous avons proposée : la mémoire collective comme partage de représentations encodées à un niveau individuel. Pourtant, nous ne pourrions nous rattacher uniquement à l'une de ces acceptions. Ces trois concepts offrent des points d'étude et seront invités à intégrer l'approche transdisciplinaire que l'histoire et les neurosciences pourront initier à l'avenir. Dans la partie suivante, nous détaillons un modèle neurobiologique simple de formation de souvenirs en illustrant les différentes dynamiques de ce processus par des études récentes.

4.1.3 Vers une neuroscience des mémoires collectives

Témoignages

L'événement commence à prendre consistance dans l'histoire à travers les traces qui s'offrent à son interprétation. Beaucoup de ce que nous savons de certaines sociétés est le fruit de travaux archéologiques, de la reconstitution à l'aide d'inférences scientifiques de l'existence de traces d'événements qui sont aujourd'hui disparus. Pour des époques plus récentes, les archives fournissent des éléments écrits permettant de rendre compte via leur interprétation de l'état de certaines sociétés ou institutions. Ces archives peuvent aussi contenir des témoignages directs, les récits rapportés de l'expérience d'un événement par un individu. Les témoignages sont des éléments cruciaux pour l'historien (Peschanski and Sion, 2018) qui va pouvoir y trouver entremêlés les descriptions de situations et leurs interprétations, leur contextualisation par les témoins directs de l'événement. L'utilisation du langage pour transmettre certaines des dimensions phénoménologiques de cet événement est la première étape de la construction mémorielle. Le langage et la mémoire épisodique sont fortement liés dans leurs évolutions respectives (Dessalles, 2007; Mahr and Csibra, 2017). Halbwachs disait : "*Nous parlons nos souvenirs avant de les évoquer*" (Halbwachs, 1925, p.279). Le langage va en effet bien au-delà d'une simple fonction d'indexation des mots aux objets, il fournit un vecteur narratif de mémoire et a épousé en partie les mécanismes séquentiels et associatifs de la mémoire épisodique pour répondre à cette demande, et la mémoire épisodique a sans doute été fortement influencée par cette nécessité de pouvoir s'accorder par le langage sur une version commune du passé (Mahr and Csibra, 2017). La capacité pour un groupe à pouvoir faire collectivement référence à un événement passé pour éclairer la situation présente a sans doute joué un rôle important dans l'évolution humaine et le développement des sociétés (Bietti et al., 2018).

Cette fonction semble d'ailleurs tellement ancrée dans la détermination des comportements que le besoin de communication et la nécessité de relater des événements vécus collectivement sont des traits qui ont été régulièrement illustrés dans les études sur le sujet. De nombreux travaux de psychologie sociale ont fait état de cette fonction régulatrice de la communication dans les jours qui suivent des événements traumatisants vécus collectivement (Rimé, 2009; Pennebaker and Gonzales, 2009). On trouve l'exemple d'une telle situation vécue et racontée par le psychologue William James lui-même dans sa correspondance avec Pierre Janet.

Alors âgé de 64 ans, le psychologue enseignait en 1906 à l'université Stanford lorsqu'un séisme frappa la ville de San Francisco. Ayant alors fait le déplacement, il remarqua dans un premier temps la rapidité et le calme avec lesquels les citoyens parvinrent à maintenir l'ordre et à porter spontanément secours aux blessés. Mais il nota aussi chez ces survivants *"un besoin irrésistible de parler de cela et d'échanger des impressions"*. A tel point que les nuits même étaient consacrées à ces échanges. *"Si on couche dehors, c'est en apparence pour être à l'abri, en réalité pour décharger avec les autres un état d'émotion et pour exprimer tout ce qu'il y avait d'inédit dans l'expérience : il y a un perpétuel bavardage toute la nuit dans les campements"* (propos rapportés par Janet, 1926). Avec cette primauté du témoignage et la prégnance de la narration dans la structuration de la mémoire épisodique apparaît en creux la possibilité d'une certaine malléabilité des souvenirs sous l'influence des interactions sociales. Ce besoin de partage est la première étape de mécanismes de convergence mnésique (Coman et al., 2016) qui va non seulement diffuser les représentations mais générer sur le long terme des représentations dominantes. L'oubli va ici revenir jouer un rôle central.

L'oubli

Ce qui intègre le contenu de la mémoire collective n'est pas la somme des représentations individuelles, non plus leur union, leurs points communs ou leur moyenne. Il s'agirait davantage de leurs prototypes, à savoir un de ces souvenirs, une de ses représentations qui va symboliser tous les autres. Si l'on considère les représentations que nous pouvons avoir de périodes historiquement distantes, nous pouvons remarquer que la variété et la diversité des récits ou des souvenirs sont souvent masquées par un faible nombre d'images ou d'informations dominantes qui sont autant de cas particuliers de souvenirs qui vont illustrer jusqu'à recouvrir la représentation complexe des événements. La Révolution française, la chute du mur de Berlin ou encore les attentats du 11 septembre à New York sont par exemple, chacun à leur façon, actuellement illustrés dans la mémoire culturelle par quelques photographies ou tableaux qui occultent l'extrême variété des représentations disponibles. Cette réduction des souvenirs paraît même d'autant plus grande que les événements sont distants dans le temps ou dans l'espace. Il semble donc que l'occultation, voire l'oubli, de certaines dimensions des souvenirs intervienne à la suite de leur grande diffusion, ou en soit même une conséquence.

Si l'oubli a été étrangement délaissé de cette approche historique et sociale de la mémoire collective (Lavabre, 2013), sa présence reste importante dans l'interprétation qu'ont livrée les sciences cognitives des mécanismes de construction de souvenirs partagés (Roediger and DeSoto, 2014; Stone et al., 2017; Brown et al., 2012). La dynamique même du témoignage est un exemple de l'importance de l'oubli dans la construction de représentations prototypiques. Ceci a été mis en évidence il y a une dizaine d'années par une série d'études menées à la *New School for Social Research* à New York et qui ont procédé à l'adaptation du paradigme d'oubli induit par la récupération de souvenirs dont nous avons parlé précédem-

ment (Anderson et al., 1994). Lors de ces épreuves, il était demandé aux participants de ré-évoquer certains souvenirs lors de discussions orales avec d'autres personnes afin de mettre en évidence une forme d'oubli induit par la récupération et partagé socialement (Cuc et al., 2007). Ainsi, lors du rappel partagé d'événements, la mémoire des personnes écoutant le récit, mais aussi la mémoire de la personne proposant le récit, étaient affectées par ce rappel, et ce même si les souvenirs étaient différents. Ce renforcement de certaines représentations se faisait aux dépens d'autres composantes, des souvenirs personnels, des détails, qui n'étaient pas évoqués lors du rappel commun. Ainsi, les mêmes mécanismes d'inhibition de la récupération, qui interviennent lors de l'oubli induit par la récupération pour empêcher l'interférence d'autres souvenirs adjacents avec l'élément rappelé, sont à l'origine de l'émergence de représentations communes gommant les spécificités de souvenirs individuels. De telles manipulations des souvenirs peuvent se faire dans des conditions de laboratoire, mais la dimension de l'événement qui est considérée et la variété des détails qui peuvent être évoqués sont alors réduites. Cette approche a été utilisée avec des récits personnels portant sur les attentats du 11 septembre 2001 à New York (Coman et al., 2009), en produisant des conclusions similaires. Dans un travail récent, (Maswood et al., 2018) ont analysé les conséquences d'un rappel collectif comparé à un rappel isolé chez des groupes d'étudiants ayant été confrontés à un événement partagé (i.e. un examen). Leurs résultats ont montré que le rappel partagé de cet événement avait réduit la charge émotionnelle négative associée à ces souvenirs, et réduisait également le nombre de détails qui étaient rapportés et augmentait la prégnance de la mémoire collective dans le sens où ce qui était mentionné par le groupe prédisait ce que l'individu allait rapporter par la suite.

Schémas mnésiques

Les étapes que nous avons mentionnées précédemment, qui vont du besoin de partage aux souvenirs partagés, conduisent à la formation des structures cognitives qui vont nous intéresser dans cette troisième étude : les schémas mnésiques (Gilboa and Marlatte, 2017; Kesteren et al., 2012). Les souvenirs sont des réseaux d'associations liant certaines composantes issues d'une expérience entre elles. Le rôle de la mémoire est de pouvoir rapporter une expérience passée lorsque quelque correspondance avec le contexte actuel l'impose. Cette nécessité est profondément ancrée dans l'évolution et n'est pas une spécificité humaine. La possibilité d'encoder la présence de nourriture ou de prédateurs sur certains lieux, et de retrouver rapidement ces souvenirs lorsque des éléments les rappellent, constitue une nécessité évidente pour les mammifères. On comprend aussi que les émotions ou les odeurs soient des attracteurs particulièrement robustes de ces associations. Là encore, l'exemple de la madeleine de Proust parle de lui-même. L'hippocampe est aussi une structure fondamentale de ce point de vue, dans le sens où sa fonction est exactement cette indexation des liens entre les détails sensoriels des expériences et leur assemblage dans le temps et dans l'espace. L'oubli, déjà, y a une influence forte car toutes ces dimensions vécues ne survivent pas une trop faible reconsolidation s'ils ne sont pas à nouveau évoqués

à l'avenir et peuvent progressivement disparaître. A l'inverse, un phénomène de consolidation et de reconsolidation (Alberini, 2005) pourra apparaître à la suite de réexpositions ou de périodes propices de sommeil (Diekelmann and Born, 2010) qui peuvent favoriser un renforcement des poids synaptiques au sein de l'engramme du souvenir ou initier des phénomènes de plasticité qui permettront de l'intégrer au sein de structures préexistantes. On considère généralement que cet ancrage plus profond des souvenirs se traduit sur le plan neural par une diminution progressive de leur rattachement à l'hippocampe : les souvenirs consolidés dépendent moins de cette indexation des expériences sensibles mais ont intégré des structures corticales pour être conservés sous une forme plus pérenne et aussi plus abstraite (Frankland and Bontempi, 2005). Bien que nous ne discutons pas ici la réalité de ce modèle, nous allons voir qu'il a pu laisser comprendre, de façon plus ou moins explicite, qu'il n'y a que deux états pour un souvenir : une version récente et vivide rattachée à l'hippocampe et une version ancienne, reconsolidée, dépendant davantage de structures corticales. Ces dix dernières années, ce modèle à deux étages s'est vu explicitement remis en cause par les études qui se sont consacrées aux schémas mnésiques. Il est notamment apparu qu'il était impossible pour un système de mémoire disposant de l'efficacité propre à celui des mammifères, de ne fonctionner qu'à l'aide de deux états (récent vs consolidé), et qu'il devait nécessairement exister un troisième système qui permettrait de lier à la fois la récupération et la consolidation des souvenirs.

L'adaptation comportementale déterminante que fournit la mémoire épisodique telle que nous la connaissons se trouve notamment **1)** dans sa grande capacité d'intégration de nouveaux éléments et **2)** dans la récupération rapide des souvenirs liés suite à l'exposition à un contexte. L'intégration de nouveaux éléments au système de mémoire à long terme est un mécanisme coûteux et risqué pour l'organisme, le risque étant la création d'interférences avec les anciens souvenirs. La récupération est complexe car toute expérience contient en elle-même une quasi infinité de relations avec des expériences précédentes. Il faut en isoler la spécificité et trouver rapidement certaines composantes de souvenirs *pertinentes* à évoquer. Les schémas mnésiques ont été proposés comme des composantes supplémentaires de ces systèmes de consolidation de la mémoire permettant d'optimiser ces comportements (Kesteren et al., 2012; Gilboa and Marlatte, 2017; Ghosh and Gilboa, 2014). Les schémas sont des structures cognitives issues de la répétition d'un ensemble d'expériences partageant des propriétés communes et dont les associations qui en structurent les éléments centraux vont être conservées au détriment des détails singuliers (Ghosh and Gilboa, 2014). L'importance de ce mécanisme a d'abord été montré chez le rat (Tse et al., 2007; Tse et al., 2011), avec le rôle du cortex préfrontal médian pour l'encodage de ces schémas qui sera également rapporté par des études en imagerie chez l'homme (voir Fig. 4.2) (Kesteren et al., 2013).

La description de ces structures cognitives préexistait pourtant avant que des études neuroscientifiques les abordent de cette façon. Le concept de schéma est apparu au début du 20^{ème} siècle (Head and Holmes, 1911) alors qu'il désignait

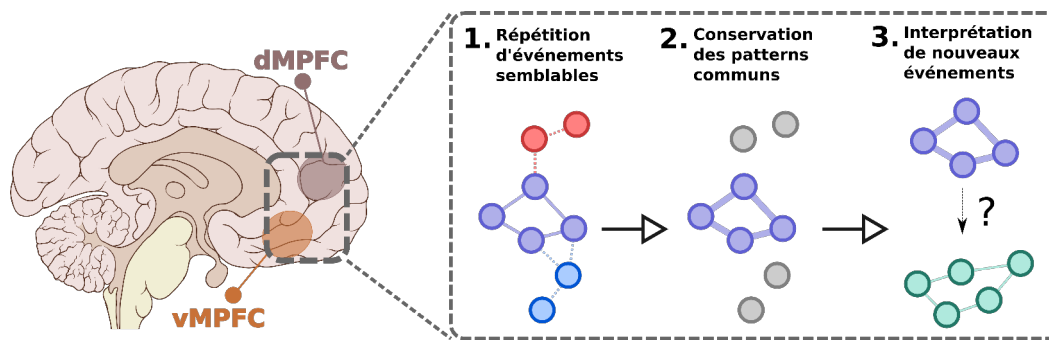


Fig. 4.2: Création et utilisation des schémas mnésiques.

pour ces auteurs “l’image corporelle”, une représentation en constante évolution de la position des membres du corps. Si cette notion est assez éloignée de l’usage que nous en avons aujourd’hui, se trouvait déjà l’idée de l’adaptation constante de ces structures et leur utilisation pour l’interprétation des événements auxquels l’organisme est soumis. Cette idée sera particulièrement reprise lorsque le psychologue britannique Frederic Bartlett en proposera une vision élaborée et assez proche de ce que nous entendons maintenant par schéma, notamment dans une dimension sociale (Bartlett, 1932/1995). Le schéma se distingue particulièrement du concept en ce sens qu’il peut intégrer constamment de nouvelles informations en fonction de la signification de nouvelles expériences (Ghosh and Gilboa, 2014). L’organisation de la mémoire humaine autour de ce type de structure a notamment pour fonction de réduire le niveau de surprise associé à une nouvelle expérience. Les modèles prédictifs de la mémoire (Henson and Gagnepain, 2010) ont participé à souligner le rôle de la minimisation des erreurs de prédiction. Si les événements ne se répètent pas à l’identique dans le temps, il est pourtant possible d’observer une certaine structure, soit par les éléments qui s’y trouvent associés ou par leur ordre de succession. Les schémas permettent ainsi, grâce à une généralisation fondée sur plusieurs occurrences d’événements semblables, d’inférer avec une plus grande précision la probabilité de co-occurrence ou de succession d’événements à partir d’observations, ce que nous pourrions rapprocher ici de l’idée d’interprétation.

Tout comme la mémoire humaine doit utiliser les souvenirs pour prédire de la façon la plus fiable possible les événements auxquels l’individu est confronté, le rôle de l’histoire n’est pas la simple conservation des souvenirs collectifs mais l’éclairage par le passé des événements présents. On comprend alors que les schémas mnésiques, ces structures cognitives encodant l’exposition répétée d’événements similaires, vont être des réceptacles particulièrement importants des cadres sociaux auxquels est exposé l’individu, que ce soit dans la communication des représentations ou à travers les ressources culturelles. A nouveau, ces associations vont être reprises pour l’interprétation des événements vécus collectivement. Il est intéressant de noter que, même en prenant des exemples très récents, l’histoire et les représentations les plus dominantes qui y figurent, sont constamment mobilisées dans la compréhension des événements collectifs. Alors que le roman d’Ernest Hemingway *Paris est une*

fête (Hemingway, 1964/2012) s'était vendu à une dizaine d'exemplaires quotidiens au cours du début de l'année 2015, ce livre est devenu, à la suite des attentats du 13 novembre 2015, un symbole de résistance symbolique et finit par figurer dans la liste des livres les plus vendus de l'année. Il en est de même pour le *Traité sur la tolérance* de Voltaire (Voltaire, 1763/1989), qui s'est 16 fois plus vendu en 2015 qu'en 2014 (Clarini, 2015), en réponse aux attaques au siège de Charlie Hebdo. Les événements historiques ont une prise rapide et profonde sur les structures de la mémoire collective préexistante par laquelle ils sont interprétés. Ainsi que le notait (Bartlett, 1932/1995), le présent ne vient jamais seul, il apparaît toujours au premier plan d'un passé structuré et cohérent par lequel il est compris.

En ce sens on peut dire que la mémoire fonde l'histoire par au moins deux de ses acceptions. La mémoire fonde l'histoire (de l'allemand *Historie*) prise comme récit. Les souvenirs sont ces composantes d'un récit qui les ordonne et les oriente selon une ligne explicative. Mais la mémoire fonde aussi l'histoire (de l'allemand *Geschichte*), prise comme la marche des événements. L'événement est à la fois une énigme dans son apparition singulière et son irréductibilité, et répétition et renaissance continuelle de son influence dans la compréhension des événements qui lui succèdent (Dosse, 2010). Si la marche des événements passés a influencé la mémoire et sa structure schématique, ces représentations préexistantes sont aussi ce qui nous permet de reconnaître l'événement comme tel : une déviation extrêmement importante de ce que nous permettait de prédire et comprendre le présent par les souvenirs. L'événement est ce point de non-retour à partir duquel nous comprenons que le futur est si radicalement changé par ce présent qu'il nous sera impossible de le comprendre sans l'intégrer dans le passé que nous prendrons avec nous pour la suite. La communication et le remodelage des souvenirs qu'elle entraîne à la suite d'un événement particulier ne se limite pas au transfert d'information, elle adapte également ces schémas mnésiques en les rendant plus cohérents avec ce que le futur réserve.

Une neurobiologie des schémas collectifs

Le travail que nous allons présenter dans cette partie concerne l'étude neuroscientifique de cette notion de schéma mnésique que nous avons étendu à une dimension sociale, culturelle et historique. Cette étude fait suite à un travail qui a été conduit en amont et qui est actuellement en cours d'évaluation (Gagnepain et al., [forthcoming](#)). Cet article a analysé l'ancrage cérébral des représentations collectives liées à la seconde guerre mondiale à l'aide d'une épreuve en imagerie fonctionnelle proposée à 24 participants après la visite du Mémorial de Caen, d'une épreuve de classification d'images proposée à 54 autres participants et de l'analyse texto-métrique de documentaires liés à la seconde guerre mondiale. Il a notamment été montré que le cortex préfrontal médian dorsal est une zone privilégiée dans l'encodage de représentations issues de la mémoire collective. L'implication effective de la partie ventrale ou dorsale du cortex préfrontal médian est un point de discussion dans les études actuelles qui ont abordé cette notion lors de tests réalisés en imagerie

fonctionnelle. De façon intéressante, la partie dorsale de cette structure est fortement associée aux compétences de théorie de l'esprit, tandis que la partie ventrale serait liée à l'extraction de patterns abstraits et serait plus impliquée dans le système de récompense. Ces résultats ont suggéré, ainsi que l'avaient proposé (Halbwachs, 1925) et (Bartlett, 1932/1995), que les souvenirs sont encodés au niveau individuel au moyen de cadres sociaux, et que ces cadres sociaux seraient fortement liés à une certaine capacité d'adopter un point de vue collectif dans l'évocation du souvenir.

4.2 Étude n°3 - Rôle du cortex préfrontal médian dans l'encodage et la récupération de représentations historiques

Les cadres théoriques qui organisent l'étude de la mémoire collective distinguent souvent deux dimensions qu'il est difficile de faire interagir. Premièrement une étude de cette mémoire par une approche "bottom-up" ou psychologisante qui stipule en particulier que les souvenirs collectifs sont des représentations encodées à un niveau individuel et conséquemment neural. Deuxièmement une approche "top-down" ou sociologisante qui intègre les représentations sociales à l'extérieur de l'individu, notamment au sein de son environnement culturel. Des interactions entre ces dimensions son souvent présupposées, mais les observations expérimentales, notamment sur le plan des neurosciences, sont encore absentes. Dans l'étude suivante, nous avons analysé en imagerie par résonance magnétique fonctionnelle les patterns d'activité liés à l'encodage et au rappel de représentations historiques chez 24 participants jeunes et ne présentant pas de troubles psychiatriques. Nous avons utilisé comme matériel des images liées à la Seconde Guerre Mondiale qui sont présentées au Mémorial de Caen. Des analyses de similarité représentationnelles nous ont permis d'évaluer la proximité de ces patterns avec des schémas collectifs que nous avons extrait de l'analyse textométrique d'un grand nombre de reportages liés à la seconde guerre mondiale et d'épreuves d'arrangement d'images proposées à d'autres participants. Cette approche permet d'évaluer l'implication de différentes zones dans l'encodage de certaines représentations. Les résultats ont montré un encodage des schémas collectifs au niveau du cortex préfrontal médian dorsal, présent lors de l'apprentissage des représentations et maintenu lors du rappel. Le titre de cet article, qui est présenté dans une version préliminaire ici est : "*Collective meaning shapes individual memories in the dorsomedial prefrontal cortex*".

Collective meaning shapes individual memories in the dorsomedial prefrontal cortex

Nicolas Legrand¹, Thomas Vallée¹, Denis Peschanski², Francis Eustache¹, and Pierre Gagnepain¹

¹Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France

²European Centre for Sociology and Political Science (Université Paris 1, EHESS and CNRS), Paris, France

Abstract

Individual memories are shaped by cultural representations. Encoding and retrieval of new experiences and knowledge are thought to be influenced by the social frameworks in which they are embedded. These frameworks can result from the repeated exposure to media, instruction, or conservation that are thought to shape our understanding of collective events. Functionally, this could engage a semantic structure or schema encoding regularities, because societies require organized knowledge. Alternatively, this could require the capability to adopt a “collective perspective”, because societies also require to understand the collective meaning of events. These questions are still unexplored in the field of neuroscience. Accordingly, while the influence of social frameworks on individual’s memory is presumed, experimental evidence is critically lacking. Here, we used fMRI to better approach the neural substrates for encoding and retrieval of information related to the Second World War (WWII). We used representational similarity analysis of patterns of activation observed in the dorsal and ventral part of the medial prefrontal cortex (mPFC) to track its consistency with collective schemas. Collective schemas were derived from the analysis of the content of thirty years of media coverage on WWII broadcast on French national television. Our results revealed a greater reactivation of collective schema in the dmPFC, but not the semantic ones, during encoding and retrieval of historical fact associated with Memorial displays. This suggests that the dmPFC might reactivate a “*collective meaning*” that shapes our memories and overcomes purely semantic frameworks. We discuss these results in the light of Halbwachs’ notion of “*social frameworks*”.

Keywords: collective memory, memory schemas, mPFC, shared representations, representational similarity analysis

Introduction

New experiences are shaped by structured and pre-existing past representations (Bartlett, 1932). This hypothesis is central to the concept of “social frameworks”, which proposes that individuals form and recall memory traces in relation to existing collective schemas (Halbwachs, 1992). However, while this collective dimension of memory has recently been further elaborated by the cognitive sciences (Hirst, Yamashiro, & Coman, 2018; Coman, Brown, Koppel, & Hirst, 2009; Barnier & Sutton, 2008; Coman, Momennejad, Drach, & Geana, 2016), the neural substrates of this influence remain largely unexplored. Recently, (Gagnepain et al., forthcoming) reported the first evidence for the similarity between collective memory schemas about World War II (WWII) events, measured from cultural resources, and the recall activity recorded in the medial prefrontal cortex (mPFC). This suggested that collective memory schemas, yet located outside individuals, can influence brain activity during the recall of items related to these particular collective representations. However, this previous study tested if individuals use collective schemas when remembering pictures from a tour at the Memorial Museum in Normandy. Thus, whether individuals rely on a set of pre-existing mental schemas encoding the structure of collective memory, or rather extracted those schemas from a Memorial representation of collective memory, remains an open question. Here, we examined using functional magnetic resonance imaging (fMRI), activity in the mPFC during both the encoding and retrieval of semantic information associated with WWII Memorial pictures and tested the underpinning influence of collective memory schema.

Collective memory refers to a set of shared representations among groups of persons (from small groups like families to larger communities like nations) that represent founding past events for the collective identity (Hirst, Yamashiro, & Coman, 2018). It is also emphasized that these shared representations may result from the social means, transmission and cultural tools (e.g. media) that a community creates (Assmann & Czaplicka, 1995). Previous works have distinguished between a bottom-up approach (how individuals cognitive mechanisms underpin the formation of shared knowledge about historical events), and a top-down approach (how cultures, mediated by several different resources, can shape and influence the cognitive structures of individuals) (Olick, 1999). These two approaches are still weakly connected in the literature, and the influence of shared representations constituting of collective frameworks found in the cultural environment on the neural organization of an individual’s

memory is debatable. The notion of memory schemas coined by (Bartlett, 1932) and later elaborated by the neuroscience (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012; Ghosh & Gilboa, 2014) provides a conceptual and neurobiological framework to account for the influence of preexisting knowledge on the encoding and retrieval of declarative memory. Here, we used an original encoding and retrieval procedure to evaluate the implication of different sub-components of the medial prefrontal cortex (mPFC) in the memorization of collective memories.

Our specific consideration for the mPFC was motivated by preexisting reports of the role of this structure in a variety of cognitive processes comprising schema abstraction, memory consolidation, decision making and social cognition (Euston, Gruber, & McNaughton, 2012; Krueger, Barbey, & Grafman, 2009). Specifically, one line of finding has emphasized the implication of the ventral part of this structure (vmPFC) in the generation of abstract cognitive structures supporting the encoding and retrieval of episodic memories through the formation of *memory schemas* (Gilboa & Marlatte, 2017; van Kesteren, Ruiter, Fernández, & Henson, 2012). Another line of evidence has reported the implication of the dorsal part (dmPFC) in the representation and mentalization of social features of events (Wagner, Kelley, Haxby, & Heatherton, 2016) and socially oriented cognitive process such as theory of mind (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). Thus, the dmPFC is often associated with perspective taking and the inference of other people’s mental states (Mar, 2011).

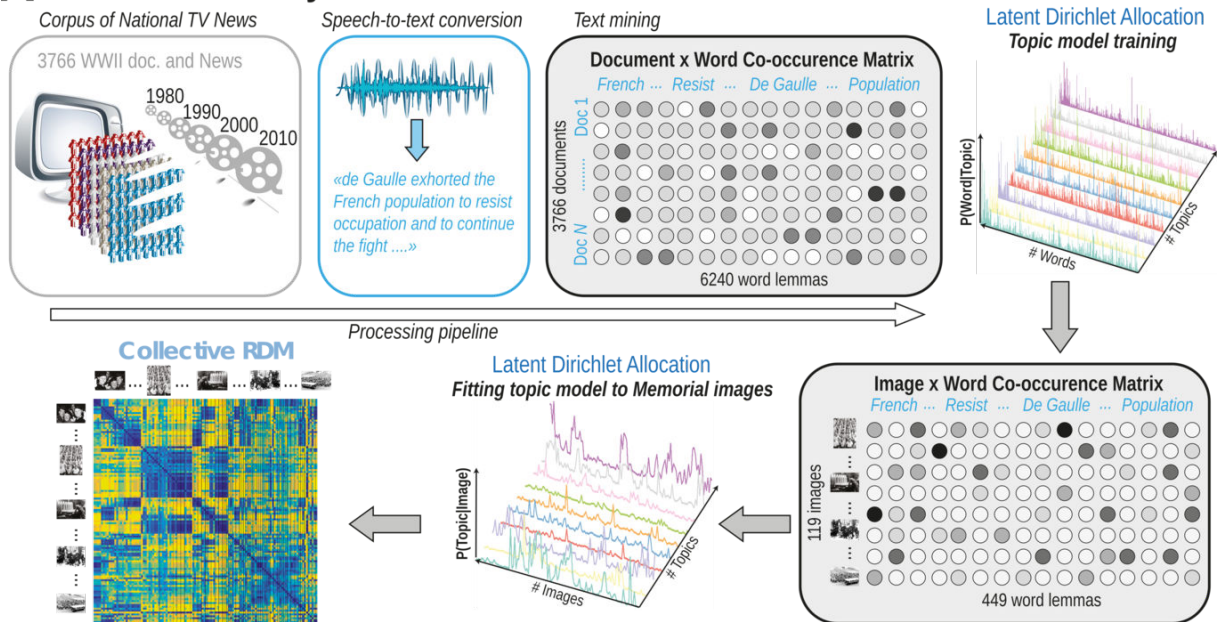
Moreover, the implication of these cortical structures to represent collective schemas during memory retrieval has recently been evidenced using a naturalistic approach (the visit of the Caen Memorial-Museum; Normandy) (Gagnepain et al., forthcoming). In this study, the organization of collective schema was capture by analyzing a corpus including all 3,766 news and documentaries on WWII broadcast on French national television from 1980 to 2010 (Fig. 1 A). This corpus was collected at the National Audiovisual Institute, an exceptional repository of all French radio and television audiovisual archives. For decades, this institute has conserved archives of audiovisual material with a particular focus on television programs, with a depository that is to be mandatory since 1992. (Gagnepain et al., forthcoming) analyze this corpus using topic modeling (Blei, Ng, & Jordan, 2003; Griffiths, Steyvers, & Tenenbaum, 2007; Steyvers & Griffiths, 2007) which is ideally suited to represent the gist or schema (Gilboa & Marlatte, 2017) of a set of words (Griffiths, Steyvers, & Tenenbaum, 2007). We summarize their approach here. After speech-to-text conversion and lexical

processing, occurrences of 6,240 canonical forms of a set of words (i.e. lemmas) were counted across all 3,766 documents to form a Word x Document frequency matrix. Second, a topic model based on a Latent Dirichlet Allocation (LDA, (Blei, Ng, & Jordan, 2003; Griffiths, Steyvers, & Tenenbaum, 2007; Steyvers & Griffiths, 2007)) algorithm was applied to the Word x Document frequency matrix. LDA uses machine learning to discover latent factors by learning the “topics” that occur in a collection of documents. LDA represents each document (here television news and documentaries) as a mixture of different topics, where a topic is a probability distribution over words. Third, this topic model was then fit to the held-out Memorial pictures using their associated captions (also lemmatized) to estimate their topic probabilities (see Fig. 1 A). The cosine distance between the distributions of topic probabilities for each pair of pictures was then encoded into a Representational Dissimilarity Matrix (RDM). This collective RDM reflects the semantic structure of the collective memory for Memorial pictures in which the meaning of a picture is defined by its location relative to all the other pictures.

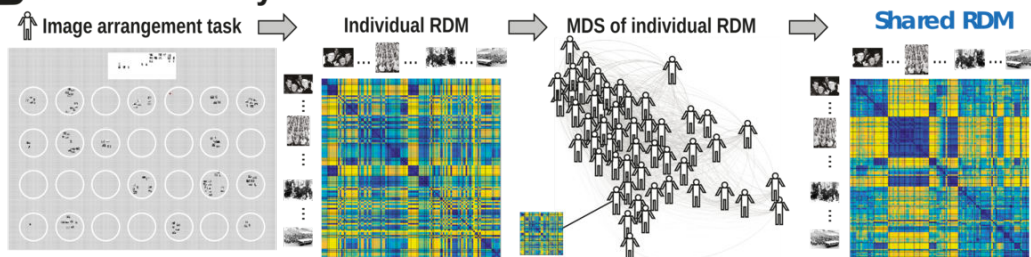
Figure 1: A. *We modeled the organization of collective memory as represented in cultural resources through the analysis of thirty years of television news and documentaries. Each audio content was transformed into text, lexically processed and analyzed using topic modeling. Then, we used the caption associated with each Memorial pictures to estimate its topic probability. The resulting distance between pictures based on the distribution of topic probabilities was encoded in a Representational Dissimilarity Matrix (RDM) and used as a measure of the semantic properties of the collective memory. The whole process of learning and fitting the topic model was repeated by varying the number of topics allowed from 1 to 100, using an increment of 1. B.* *The optimal number of topics was then determined using an image arrangement task proposed to 54 individuals unfamiliar with the Memorial. Participants had to arrange a set of images from the museum based on their historical proximity. We then computed a shared RDM corresponding to the compromise between all 54 individual RDMs derived from the Euclidean distances of the spatial arrangement. This procedure showed that collective and shared memories were extremely similar across control individuals (see Figure 3A), and reached its maximum when 6 to 10 topics were included during topic discovery. C.* *Brain RDMs were extracted in the two regions of interest (dmPFC and vmPFC) and reflected the correlation distance (1- correlation) between the 63 pictures used. These RDMs were then compared to the RDM representing the collective memory.*

Our goal was to estimate, both during memory encoding and retrieval, the representational similarity (Nili et al., 2014) between the organization of collective RDMs on one hand, and

A Collective memory schemas



B Shared memory schemas



C Individual memory organization

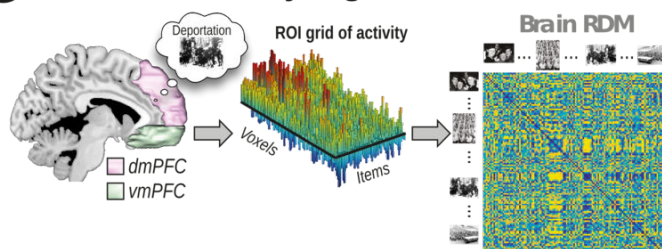


Figure 1: Measures of collective, shared and individual representations.

the organization of individual memory representation in the vmPFC and dmPFC, on the other hand. For this, we included 24 participants in an fMRI procedure that comprised an encoding and retrieval phase of Memorial displays related to WWII (see Figure ?? and Method section for details on the procedure). The encoding phase consisted in 3 repeated

presentations of 63 pictures originally exposed inside the Caen Memorial. The picture was described by a very short caption and each picture presentation was followed by a question concerning the location (“Where?”), the date (“When?”) and the content (“What?”) of the scenes (questions order were randomized across pictures and participants). The correct response was then displayed to the participant who was encouraged to use this feedback to increase his/her knowledge of the picture. In a subsequent phase realized the next day, we first assessed their declarative memory for semantic facts associated with the pictures and learned the day before, this time without feedback.

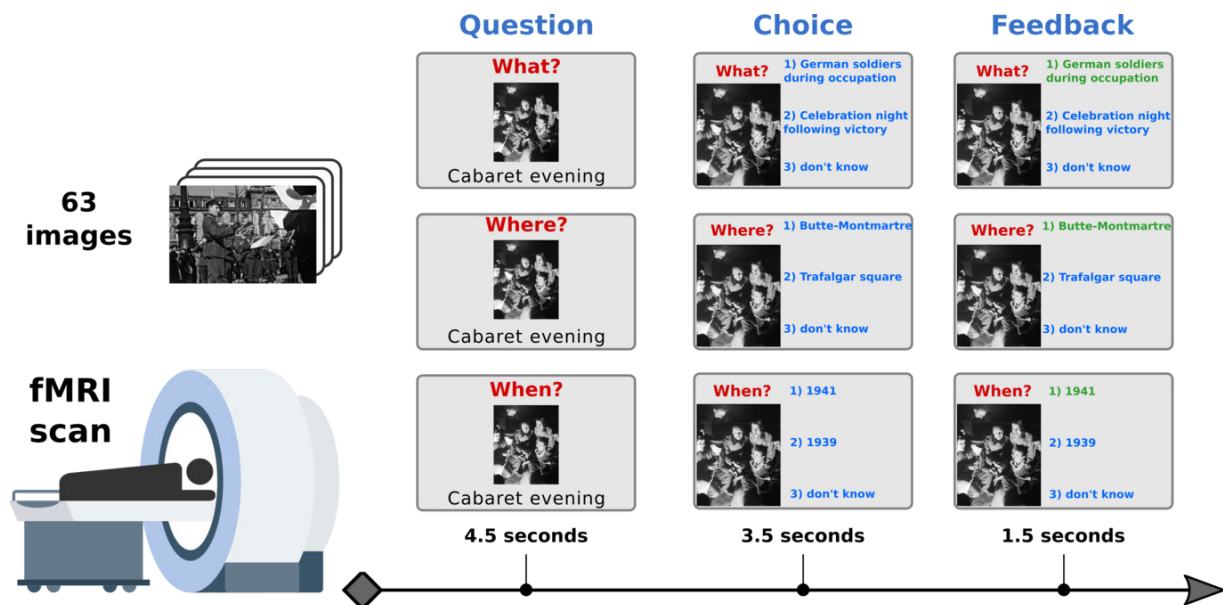


Figure 2: **Experimental design to measure declarative memory.**

Figure 2 : During day 1, they viewed 63 images representing events from the WWII. Each picture was presented 3 times and followed by a question: “where?”, “what?” and “when?”. The image and the main questions were presented for 4.5 seconds. Then, three choices were provided to the participants: two were potential correct responses, among which one was systematically the correct, and a third choice if the participant did not know the correct answer. They were asked to provide a correct response without guessing, and to use the third choice otherwise. During day 2, the task employed was similar to the first initial encoding task and evaluated their declarative memory of the semantic information associated with the image. However, no feedback was given after participant’s response during day 2.

Controlling for shared semantic

Furthermore, the relationship between pictures captures by our model of collective memory could correspond to pre-existing semantic similarities also reflecting a form of shared (although not collective) memory for common concepts and the meaning of language. To control for this commonly shared semantic relationships, we used 2643 French Wikipedia articles related to WWII as a benchmark model of general semantic relationships between words and trained a topic model that we then fit to the Memorial pictures. LDA models of Wikipedia articles about various concepts have been shown to produce an accurate representation of semantic features that predict well patterns of brain activity (Pereira, Detre, & Botvinick, 2011; Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). As a place of global remembrance where the memory is not simply stored but is the product of a collaborative recall between many people and has a communicative function, Wikipedia shares some features with the concept of collective memory. There are, however, few critical distinctions. Collective memory is a selective representation of the past, shaped by schematic narrative templates that contributes to the construction of the group identity, often emphasizing some elements while minimizing others. Given this encyclopaedical nature, it is debatable if the Wikipedia does foster this selective representation of past and the formation and compilation of corresponding memories that bear on the group identity. Wikipedia cannot be understood as one consistent medium like television which consistently promotes symbolic and memorable elements of collective memory without negotiation. For instance, the meaning of the word “collaboration” will be really different when comparing the collective TV News corpus and Wikipedia. This concept unambiguously refers to the participation of French state and its politic, and their representative at the time (e.g. Bousquet, Leguay, Touvier, Papon), to Jew deportation and destruction, or to help the German Nazi arresting French resistant. In WWII-related Wikipedia articles, the word “collaboration” is associated with very different semantic contexts not necessarily related to French collaboration. Thus, Wikipedia RDMs were created using exactly the same method developed to create collective RDM, and served as a control for shared semantic meaning.

Validation of collective memory RDM

We then ran a series of additional comparisons to validate our model of collective memory derived from the corpus of national News, and demonstrate the universality of the collective memory RDM derived from topic modeling. Our goal was to demonstrate that the collective

memory RDM corresponded to a shared and selective representation of the past, whose core structure is different from a shared semantic meaning (i.e. Wikipedia RDM). Our perspective thus treats collective memory as located both in social tools and means used to connect individuals (such as national News) and in individual minds themselves, who thus share this memory as events whose meaning and social value is important to the society. We sought to assess how well the previously generated collective structure matched a shared pool of knowledge across 54 control individuals unfamiliar with the Memorial, and whether this shared representation was a better characterization of the collective memory RDM than an RDM describing image relatedness based on shared semantics. To capture the shared structure of individual representations, we designed an image arrangement task (Kriegeskorte & Mur, 2012) (Fig. 1 B) aiming to capture each individual semantic organization of Memorial pictures within a common space. During this task, we asked participants to position the images that are exposed at the Memorial according to their historical proximity within the circles partitioning the map space (Fig. 1 B). Those spatial arrangements of historical pictures reflect the semantic organization of a given individual and can also be encoded within an RDM using the Euclidean distances between images. To measure a common representation reflecting a shared schema (i.e. shared RDM), we computed the compromise (Abdi, Williams, Valentin, & Bannani-Dosse, 2012) of those 54 individual RDMs. Collective RDM extracted from the News corpus was extremely similar to the structure of the shared RDM measured across control individuals (Fig. 1 B). This similarity between collective and shared memory reached its maximum when a total number of 6-10 topics were included during topic discovery. Furthermore, the collective RDM was best predicted by a shared RDM which outperformed Wikipedia RDMs. As a result, all subsequent analyses were based on a collective memory RDM computed using 6-10 topics.

Results

During dmPFC encoding, we found a near significant relationship between activity patterns and collective schema ($P = 0.052$; 95% Confidence Interval (CI) = [-0.0028 0.0303]), while no significant relationship was found for semantic schema ($P = 0.33$; 95% CI = [-0.0124 0.0180]). During vmPFC encoding, we showed a consistent relationship between brain RDM and both collective ($P = 0.0008$; 95% CI = [0.0094 0.0505]) and semantic ($P = 0.0008$; 95% CI = [0.0058 0.0589]) schemas. During encoding, a near significant

difference were observed between collective and semantic RDMs for the dmPFC ($P = 0.08$; 95% CI = [-0.0041 0.0238]) and no difference were observed for the vmPFC ($P = 0.42$; 95% CI = [-0.0204 0.0232]). The proximity between the brain RDMs and the collective and semantic RDM thus indicate that the neural representations in the vmPFC code for abstract knowledge irrespective of their collective memory status. On the opposite, the dmPFC seems to preferentially respond to the influence of collective schema. This was confirmed during the analysis of the retrieval phase on day 2. During dmPFC retrieval, we found a significant relationship between brain RDM and collective schema ($P = 0.0234$; 95% CI = [0.0007 0.0283]). This relationship was not observed for semantic schema ($P = 0.4936$; 95% CI = [-0.0145 0.0141]) schemas. Critically, during retrieval, the similarity of the dmPFC representational content with collective RDM was stronger than the similarity with semantic RDM ($P = 0.0020$; 95% CI = [0.0047 0.0241]). The same difference was found for the vmPFC activity ($P = 0.0052$; 95% CI = [0.0025 0.0222]), but this time the collective schema did not show a significant relationship with vmPFC RDM ($P = 0.4742$; 95% CI = [-0.0124 0.0141]) and the semantic schema exhibited an inverse relationship ($P = 0.0344$; 95% CI = [-0.0231 0.0005]), evidencing for a greater distance with the semantic schema. Finally, when contrasting the main effect of collective and semantic schemas across encoding and retrieval phases, we observed a significant difference in the dmPFC ($P = 0.0018$; 95% CI = [0.0038 0.0203]) but not in the vmPFC ($P = 0.14$; 95% CI = [-0.0052 0.0191]). These findings then suggest that the patterns of activity in dmPFC is preferentially influenced by collective schema, while the influence of abstract schema (irrespective of their collective status) is mostly observed in the vmPFC during the learning of semantic-picture associations. We then thought to assess whether the reactivation of collective schema (contrasting against semantic schema and collapsing across encoding and retrieval) was predictive of an increase in memory performance between day 1 and day 2. This increase in memory performance ($M = 21\%$; $STD = 10$) was computed as the difference between the proportion of correct response to the What/When/Where questions during encoding (day 1) and retrieval (day 2). Using robust correlation (Pernet, Wilcox, & Rousselet, 2013), we found a significant relationship between this increase in memory performance and the reactivation of collective schema in the dmPFC ($r\text{-spearman} = 0.5562$, 95% CI = [0.0940 0.8285]; 6 bivariate outliers automatically detected). Note that this pattern was also observed separately for encoding and retrieval.

Figure 3 : Panel A displays the correlation between shared RDM and collective/Wikipedia RDM (averaged by range of 5 topics). This similarity between collective and shared memory

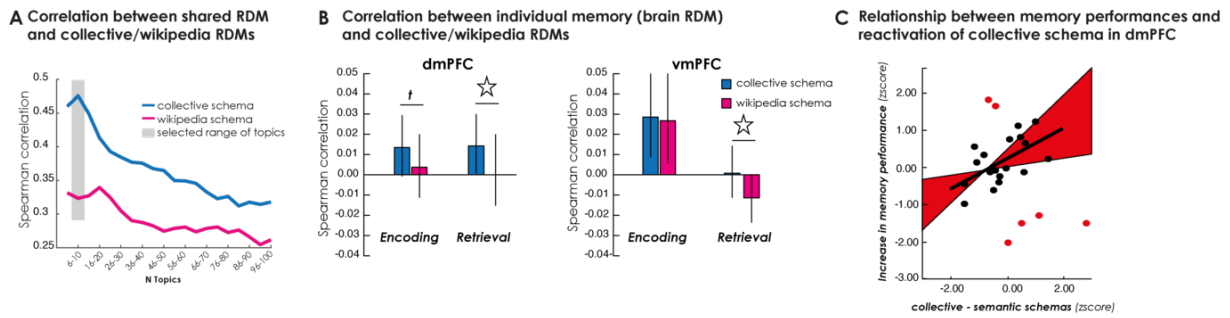


Figure 3: Correlation between shared and individual memory (brain RDMs), and collective and Wikipedia RDMs.

(blue line) was much higher than the similarity with Wikipedia RDM (pink line) and reached its maximum when a total of 6-10 topics were included during topic discovery. Panel B displays the average correlation across participants ($n = 24$) between the upper triangular forms of dmPFC (left panel), vmPFC (right panel) RDMs, and collective and Wikipedia model RDMs, during encoding and retrieval phases. Error bars reflect 95 % bootstrapped confidence intervals and thus reveals a significant relationship between models and brain RDMs when they do not overlap with zero. Stars and "t" indicate significant differences and trend, respectively. Panel C displays the significant relationship between the increase in memory performance (day 2 - day 1) and the reactivation of collective schema (versus semantic schema) in the dmPFC.

Discussion

Almost a hundred years ago, it was hypothesized that memory reconstructs past experiences based on social frameworks bound to collective memory. However, the notion that such collective schema would be encoded in the brain in the form of preexisting knowledge is still lacking experimental support. This question has been recently addressed by cognitive and social psychology (Pennebaker & Gonzales, 2009; Hirst, Yamashiro, & Coman, 2018; Coman, Brown, Koppel, & Hirst, 2009), but neuroscientific evidence is still critically lacking. Here, we attempted to assess the influence of collective schemas, extracted from the content of thirty years of media coverage on WWII broadcast on French national television, on the pattern mPFC activity during both learning and retrieval of semantic information associated

with WWII Memorial pictures.

Our results revealed that collective and semantic schemas modulate activity in the dmPFC both during the encoding (day 1) and retrieval (day 2) of historical fact associated with the Memorial pictures. This is consistent with a previous report using the same collective schema in a different setting (Gagnepain et al., forthcoming). Importantly, this reinstatement of collective schemas is not reducible to a semantic architecture derived from an analysis of WWII Wikipedia articles. A collective schema is therefore not comparable to an organized set of abstract knowledge but might imply distinct representations that we refer here as a “collective mental map” or a “collective perspective”. The dmPFC is involved with processes related to the theory of mind or the representation of other’s point of view (Mar, 2011; Wagner, Kelley, Haxby, & Heatherton, 2016; Overwalle, 2009; Wagner, Haxby, & Heatherton, 2012). Critically, activity in the dmPFC has been linked to the representation of shared events across individuals (Chen et al., 2016), supporting the notion of a neurocognitive marker of the collective viewpoint. This interpretation is also supported by the close similarity we observed between collective schema and shared memory representation (Figure 3 A). Our findings might thus reflect the need for collective schemas to represent a perspective that transcends the individuals. Adopting such a “*collective perspective*“, or “*social framework*” might be necessary to correctly encode and retrieve declarative memories about historical events, as illustrated by the relationship observed with memory performances in Figure 3C. Most importantly, this suggests that collective memory, which exists outside and beyond individuals, might constitute a common mental model connecting people’s memories across time and space. This supports the notion that this area is functionally engaged during in the elaboration and reconstruction of memories associated with collective representations. On the opposite, both semantic and collective schemas were found to influence patterns of activity in vmPFC, but mostly during encoding. This is in line with previous works showing that this structure is involved in the formation of abstract structures encoding the recurrence of associations like memory schema (van Kesteren, Ruitter, Fernández, & Henson, 2012; van Kesteren, Fernandez, Norris, & Hermans, 2010; Spalding, Jones, Duff, Tranel, & Warren, 2015; van Kesteren et al., 2013).

Collective memory is a selective representation of the past that is continuously reconstructed to fit with the actual course of collective events and the need for understanding collective points of view. Individual memories are also shaped by these contextual social frame-

works (Halbwachs, 1925/1994; Halbwachs, 1992). Halbwachs famously exemplified this notion with representative states of altered remembering like dreaming, where individuals cannot truly retrieve memories due to their disconnection not only with the external world but also with the social framework providing the central concepts with which memories are restructured. Indeed, memory corresponds to reconstructive processes (Bartlett, 1932) and arises from the reactivation of very distinct neocortical components (Rolls, 2013). Many factors are thought to influence or interfere with this reconstructive process. Here, we specifically addressed the question of "social networks" (Halbwachs, 1925/1994; Halbwachs, 1992), which presumes that individual representations are shaped by the cultural, historical and interpersonal context of remembering. The so-called "social turn" operated in memory researches (Hirst & Rajaram, 2014), and the recent developments in psychology around the cognitive processes supporting the creation and emergence of shared representations (Coman, Momennejad, Drach, & Geana, 2016), have corroborated this notion. However, the connection between the social and cultural inscriptions of memories, and its inscription to the neural engram of memory is an epistemological and methodological challenge. The ontology of collective memories (i.e. whether it can or not be conceived outside the neural system) is also debated (Olick, 1999; Halbwachs, 1925/1994).

Here, we grounded our approach on the role of memory schemas as the articulation of reconstructive remembering, both from a psychological and cultural perspective (Wagoner, 2013). This supposes that schemas, which extract recurrent patterns of complex associations that individuals used to navigate social and cultural interactions, can be found in the environment they arose from and in the cognitive system of the individuals. Using RSA (Nili et al., 2014; Kriegeskorte, 2008), we show that these structures could be mathematically described and then apprehended from both perspectives. RSA constitutes, therefore, a central approach to assess the interaction between distant scientific representations, and is in itself constitutive of such trans-disciplinary approaches.

Method

MRI Participants. Twenty-four right-handed native French speakers between the ages of 20 to 33 years ($M = 25.9$; $SD = 3.3$) were paid to participate (13 males). They had no re-

ported history of neurological, medical, visual or memory disorders. The study was approved by the regional research ethics committee (*Comité de Protection des Personnes Nord-Ouest III*). All participants gave written consent before participation and were unfamiliar with the Caen Memorial. Participants were asked not to consume psychostimulants, drugs, or alcohol prior to or during the experimental period.

Material. The stimuli were 63 pictures selected from the area “World War, Total War” at the Caen Memorial-Museum (Normandy). Each display was photographed at the Memorial using professional numerical camera and lighting. 63 additional high-quality pictures, picked up from various WWII historical website on the internet, were selected as the foils and used for the recognition task during day 2. Both sets of pictures of pictures were then adjusted for contrast and lighting, and external contour was cropped, using an image processing software.

Encoding/Retrieval task. Participants were first asked to perform an encoding task inside the MRI scan on day 1. This task was divided into 5 fMRI sessions, each about 7 minutes in length. During this procedure, the 63 pictures were presented on the screen for 4.5 seconds, accompanied by a few words captions describing the picture, and one of the following question: “What?”, “Where?” or “When?”. Thus, each picture was presented 3 times (with a different question each time), leading to a total of 189 stimuli presented in randomized order. After 4.5 seconds, the picture remained on the screen but three choices appeared on its right for 2.5 seconds: two sentences described potential correct responses (one right and one wrong) and an additional “don’t know” response. We asked the participant to answer as accurately as possible without trying to guess and to select the third option in case they did not know the correct answer. After they selected one of the proposition, a feedback was provided and the correct answer appeared in a green color. Participants were encouraged to use this feedback to increase their knowledge of the picture. On day 2, the participants performed the exact same task, this time without feedback. The task was divided into 4 fMRI sessions, each about 8 minutes in length.

Recognition task. Just before the retrieval task on day 2, participants performed a recognition task, during which we presented in randomized order the 63 pictures that the participant had learned the day before, presented along with 63 foils pictures with similar characteristics. The participant had to report whether he/she saw or not the picture during

day 1. The recognition task was performed to ensure that pictures included in the RSA were correctly recognized (not recognized pictures were not included).

MRI acquisition parameters. MRI data were acquired on a 3T Achieva scanner (Philips). All participants first underwent on day 1 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, slice thickness = 1 mm, no gap, FoV = 256×256 ×180 mm³, matrix = 256×130×180). The functional sessions 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 64×63×32, FoV = 192×192×119 mm³, 310 volumes per run).

MRI preprocessing. Data were analyzed using Statistical Parametric Mapping software (SPM12, Wellcome Department of Imaging Neuroscience, London, UK). During preprocessing, images were first spatially realigned to correct for motion, before being corrected for slice acquisition temporal delay. After coregistration with the T1 structural image, functional images were then normalized using the parameters derived from the nonlinear normalization of individual gray-matter T1 images to the T1 template of the Montreal Neurological Institute (MNI). Note, however, that unwarped and unsmoothed images were used for RSA. Image normalization was nonetheless needed to compute the forward deformation field and its inversion, to wrap back mPFC ROIs to native space (see below). The use of unsmoothed images is important for RSA as it preserves the fine-grained spatial pattern that characterizes the representational geometry of a region.

First level analysis and Representational Similarity Analysis (RSA). The pre-processed time series, corresponding to native space images (i.e. non-warped and unsmoothed images), were then high-pass filtered to 1/128 Hz in each voxel. Regressors within a general linear model (GLM) for each voxel were created by convolving a delta function (modeled as 6 seconds short-epoch) at stimulus onset for each condition of interest with a canonical HRF. Regressors of no interest were the six realignment parameters to account for linear residual motion artifacts. Autocorrelation between the GLM residuals was corrected using the first order autoregressive process resulting in pre-whitened data after restricted maximum likelihood estimation. A “Least-Squares Separate” approach was used (Turner, Mumford, Poldrack, & Ashby, 2012; Mumford, Turner, Ashby, & Poldrack, 2012) which con-

sisted in estimating a separate GLM for each trial. Within each GLM, the trial of interest is modeled as one regressor, and all the other trials are collapsed into a distinct regressor. This approach has been promoted for designs with short ISI, when there is a high level of collinearity between BOLD responses to successive trials (Mumford, Turner, Ashby, & Poldrack, 2012). However, our goal was to use the three repeated presentations of each item (corresponding to the What/Where/When question) to compute a cross-validated estimate of the similarity between each pair of pictures for the purpose of RSA (Walther et al., 2016). To this end, we repeated 3 times the “Least-Squares Separate” approach for each item. For each iteration, we created three distinct regressors: the first “training” regressor combined two repetitions of an item (e.g. combining What and When questions), the second “testing” regressor encoded the remaining presentation of this item, and the third one, all of the other items. This procedure was then repeated for each item, producing for each voxel and each item, 3 “training” and “testing” parameters estimates (corresponding to all permutation orders of the 3 questions) that will be used to compute cross-validated similarity. RSA allows a comparison of different representational models (e.g. collective schema in the current study) which can be defined as hypotheses about the distance or similarity of activity profiles across experimental conditions. Noise can, however, create a positive bias in the distance measurement (Walther et al., 2016). The cross-validated distance between repeated stimuli should be however unbiased by noise, and represent the expected true distance with a zero-interpretable point indicating the absence of differences between patterns. Here, we compute the similarity for all pairs of possible pictures by computing the dot product of the training and testing vector of activity and averaging these cross-validated similarity values across the 3 permutations of our cross-validation scheme (and then shift the sign to store them as representational dissimilarity matrix). The upper triangular form of these brain RDMs (i.e. see ROI definition below) was then extracted and compared to the upper triangular form of RDM models describing collective and semantic (i.e. Wikipedia) schemas. These comparisons between brain RDMs and a given model were achieved using the Spearman correlation coefficient. For both the encoding and the retrieval phases, only pictures with at least one correct response across the 3 questions were included. Moreover, only items correctly recognized during day 2 were included in the analysis of activity patterns of both the encoding and retrieval phase. Group-level inferences were conducted using nonparametric random effects statistics to test for both RDM relatedness and differences by bootstrapping the subject set with 5,000 iterations (Nili et al., 2014). For each model RDM or each pair of contrasted model RDMs, we performed a two-sided statistical comparison and estimated the

p-value as the proportion of bootstrap samples further in the tails than 0 (Nili et al., 2014). We also report and use bootstrapping iterations to determine 95% confidence intervals.

ROI definition. The vmPFC and dmPFC ROIs were defined anatomically using the AAL atlas (Tzourio-Mazoyer et al., 2002) and were similar to those defined by (Gagnepain et al., forthcoming). The dmPFC corresponded to the bilateral frontal superior medial gyrus of this atlas (indexes 2,601 and 2,602). The vmPFC mask included the bilateral frontoorbital medial gyrus (indexes 2,611 and 2,612), the bilateral rectus (indexes 2,701 and 2,702), as well as the ventral portion (Z coordinates inferior or equal to zero) of the bilateral anterior cingulum (indexes 4,001 and 4,002). These two mask images were then wrapped back to each participant native space using the inverse of the deformation field computed during the normalization process.

Collective memory corpus description and analysis. The corpus was collected by the MATRICE project (<http://www.matricememory.fr/?lang=en>), a multidisciplinary and technological platform, whose aims is to provide tools, technological and theoretical background to understand the relationship between collective and individual memory. The audio-video clips initially composing the corpus were stored at the « *Institut National de l'Audiovisuel* » (INA). INA is one of the main partners of the MATRICE project and a public institute whose one of the main goals is to archive all audiovisual productions broadcast on French television or radio. For this particular study, we included in our corpus all television news and documentaries (not including radio shows) broadcast from 1980 to 2010 with WWII for common theme, leading to a total of 3,766 documents. We focus on this particular time period for three reasons. Firstly, 1980 to 2010 largely overlap with the live period of our participants. Secondly, this period also corresponds to the establishment of a new narrative for French collective memory (i.e. a new “*Régime of Mémorialité*”, see (Peschanski, 2012)). This included emergence and affirmation of the Shoah memory, as well as major trials acknowledging the participation of French state and its politic, and their representative at the time (e.g. Bousquet, Leguay, Touvier, Papon), to Jew deportation and destruction. Third, thanks to advances in automatic speech recognition and to the availability of electronic texts with which contemporaneous language models were built to process data recorded after 1980, the 3,776 audio files were converted to XML format using speech-to-text conversion algorithms developed by the LIMSI (« *Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur* » / Engineering Sciences and Compu-

ter Science lab), one of the main partners of the MATRICE project. Once converted to text, our corpus was initially manually processed to discard segments not related to WWII (exclusively keeping sections of the news or documentary dedicated to this topic). During this operation, the automatic speech-to-text transcription was additionally corrected by a human reader. Following this initial preprocessing, a separate XML file was extracted for each document. These files were then processed with textual data analysis methodologies using the TXM software (see (Heiden, 2010); <http://textometrie.ens-lyon.fr/>) connected with TreeTagger morphosyntactic analyzer (see (Schmid, 1994); <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>). TXM allows to annotate each instance of a word (i.e. token) to its corresponding lemma (a canonical form of a set of words) and can proceed to a morphosyntactic labeling of each word in a corpus. The corpus can then be explored using Corpus Query Language (CQL) requests to tag and retrieve specific grammatical categories, forms, lemmas, or any information attached to the word. A list of CQL request was first automatically generated for each noun, verb, or adjective lemma that could find in this corpus. Rare lemmas (less than 10 occurrences), stop words, and common verbs (e.g. “be”, “have”) were removed in the process and not included in subsequent analyses. For instance, this algorithm would go through the corpus and produce a unique CQL query, e.g. [lemma=“fight”%c], corresponding to any instance of the verb “fight” in the corpus (i.e. “fight”, “fought”, “fighting”). The %c modifier incorporated in this request is used to neutralize the character case of assimilated lemma forms (i.e. “fight”, “Fight”, “FIGHT”). An initial list of about 6500 CQL requests was automatically generated this way. This list was manually checked to remove CQL requests produced by wrong transcription, to occasionally group certain requests under a single entity (e.g. [lemma=“drama”%c][lemma=“dramatic”%c]), to merge bigrams when appropriate (e.g. [lemma=“chief”%c][lemma=“state”%c] for “chief of state”), or to adjust to different orthographical forms (e.g. [lemma=“Gorbachev”%c][lemma=“Gorbachov”%c]). Following this initial check and adjustment, a total of 6,240 requests remained to summarize the 1,431,790 tokens composing our corpus.

A lemma x document frequency matrix counting the number of occurrences was extracted and submitted to a Latent Dirichlet Allocation (LDA) analysis performed using the Machine Learning for Language Toolkit (MALLET; <http://mallet.cs.umass.edu/topics.php>). Topic models (Blei, Ng, & Jordan, 2003; Griffiths, Steyvers, & Tenenbaum, 2007; Steyvers & Griffiths, 2007) consider that documents are generated by mixtures of topics. In this con-

text, a topic corresponds to the distribution of probabilities over all words present across all documents (how likely is a given word associated to a topic). From this perspective, a document can be generated by assigning a probability distribution over topics. For each instance in a document, a topic is chosen depending on its previous topic probability distribution and a word is drawn from that topic. MALLET uses Gibbs sampling algorithm to invert this process, inferring the set of topics that were responsible for generating a collection of documents and their probabilities over words.

We first trained topic models on the WWII news and documentaries corpus using the MALLET train-topic engine. We learned the topic models varying the number of topics allowed from 1 to 100, in increments of 1, setting the alpha parameter to $50/N$ topics and the starting Beta parameter to 0.1 (as suggested in other works modeling a large corpus of texts for semantic purposes, see (Griffiths & Steyvers, 2004)). For each number of topics, we used 500 iterations to estimate word and document topic probabilities. Once estimated, we used the MALLET inferencer tool to fit the LDA model to the held-out Memorial pictures and estimate their topic probabilities. For that purpose, each image was treated as a new document and labeled with keywords (also lemmatized) which were directly derived from the Memorial captions underneath the exposed pictures. A total of 449 lemmas was used to describe the Memorial pictures. All of these lemmas were also found in the list 6240 lemmas describing the corpus of news and documentaries. This inferential process thus led to a probability distribution matrix of 63 images x N Topics, describing the posterior probability of a topic given an image.

A 63 images x 63 images RDM was then computed for each number of estimated topics using the distance between the distributions of topic probabilities for each pair of pictures (based here on the cosine distance). However, given the randomness that the Gibbs sampling algorithm may introduce during parameter estimation, we reiterated the whole process 10 times leading to a 3-Dimensional 63 images x 63 images x 10 repetitions RDM for each number of estimated topic. To summarize these RDMs across the 10 repetitions of the topic modeling estimation, we used DISTATIS which is ideally suited to compute the best agreement or compromise across multiple distances matrices (Abdi, Williams, Valentin, & Bannani-Dosse, 2012). The implementation in Matlab of DISTATIS can be found here: <https://www.utdallas.edu/~herve/>, but in brief: **1)** each of the 10 RDMs are first transformed to a cross-product matrix after double-centering and normalization to their first eigenvalue, **2)** the cosine similarity structure of all pairwise comparisons of the 10 normalized cross-product matrices is computed using the Rv coefficient, **3)** the Rv coefficient matrix

describing the relationship between RDMs is submitted to eigen-decomposition, and the compromise matrix corresponds to the sum of normalized cross-product matrices weighted by their first eigenvalue, **4**) the eigen-decomposition of the compromise produces factor scores which describes the position of each of the 63 images in a N-dimensional compromise space, and **5**) the distance correlation matrix of these multidimensional compromise space then correspond to the best agreement across all 10 collective RDM derived from the repetition of topic estimation.

To summarize the whole process, for a given number of topics: **1**) we first train a topic model on the French television news and documentaries corpus, **2**) then we fit this topic model to the Memorial pictures and their captions, treating each picture as a new document, **3**) a 63 pictures x N Topics matrix, describing the posterior probability of a topic given an image, is extracted and transformed to 63 x 63 RDM, **4**) this process is reiterated 10 times and the best agreement across the 10 RDMs produced at each repetition is computed. These steps are repeated for each number of estimated topic, thus producing a 3-Dimensional 63 pictures x 63 pictures x 100 topics RDMs.

Validation of collective memory model. Many scholars treat collective memories as shared individual memories or representations that result from social means and transmission tools that a community creates. These social means not only have the function to reshape and harmonize individual representations by guiding the remembering of all members of the society but also regulate the construction of a collective identity (Olick, 1999). This perspective thus treats collective memory as located both in social tools and means used to connect individuals (such as national news) and in individual minds themselves, who thus share this memory as events whose meaning and social value is important to the society. To validate our model of collective memory derived from the corpus of national news and documentaries, and demonstrate the universality of the collective memory RDM derived from topic modeling, we sought to quantify the structure of shared representations across individuals and compared such shared schema to our collective memory model. Furthermore, this procedure would also permit to select independently an optimal number of topics to describe the collective structure of pictures which best corresponds to shared memory.

To this end, 54 native French speakers between the ages of 20 to 39 years ($M = 27.3$; $SD = 5.6$) performed (23 males) the image arrangement task (see the supplementary section dedicated to this task above). This task was performed on the exact same 63 Memorial

pictures but unlike our MRI participants who visited the Memorial the day before, those new participants were completely unfamiliar with the Memorial. 54 individual RDMs were thus derived from the Euclidean distances between images. We performed DISTATIS (*Abdi, Williams, Valentin, & Bennani-Dosse, 2012*); see above) to capture the shared structure of representations across those control individuals. Indeed, the distance correlation matrix of this RDMs compromise corresponds to the best agreement across all 54 participants and thus can be viewed as a shared schema reflecting the common semantic organization across individuals. Collective RDMs extracted from the corpus of television news and documentaries were extremely similar to the structure of shared memory measured across control individuals. This similarity between collective and shared memory reached its maximum when an average total of 6-10 topics was included during topic discovery. As a result, all subsequent analyses involving collective memory were performed using 6-10 topics.

Construction of a control model of verbalizable semantics. The relationship between pictures captured by our model of collective memory could reflect pre-existing semantic connections, which also exist outside and transcend individuals, and also reflect a form of shared (although not collective) memory for common concepts. It is therefore critical to show that the performance of the LDA model in predicting brain activity depends on the fact that it was specifically trained on a corpus of French TV news specifically collected to reflect collective memory. If an LDA model trained on a general corpus which also contains commonly shared semantics serves as a strong RDM predictor of mPFC dissimilarities, this would indicate that our collective memory model has in fact captured semantic relationships irrespective of the collective content of memory. If, on the opposite, our LDA model of collective memory outperforms an LDA model trained on a more general corpus, this would be a strong indication that the mPFC is indeed encoding a common mental model reflecting the structure of collective memory.

LDA models of Wikipedia articles about various concepts have been shown to produce an accurate representation of semantic features that predict well patterns of brain activity (*Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016; Pereira, Detre, & Botvinick, 2011*). We use the whole French Wikipedia articles as a benchmark model for general semantic relationship between words. This corpus (<http://redac.univ-tlse2.fr/corpus/wikipedia.html>), included 2643 articles related to WWII and edited until June 2008. These articles have been previously extracted from the dump French version of Wikipedia articles (<http://dumps.wikimedia.org/>) and processed using the same morphosyntactic tagging

tool (Schmid, 1994)(Schmid, 1994) used to process our corpus of French TV News on WWII. Once imported into TXM, we applied exactly the same method of analysis that we previously used to construct the model of collective memory (cf. *Collective memory corpus description and analysis* section). To summarize the whole process again, for a given number of topics: 1) we first train a topic model on the whole French Wikipedia corpus, 2) then we fit this topic model to the Memorial pictures and their captions, treating each picture as a new document, 3) a 63 pictures x N Topics matrix, describing the posterior probability of a topic given an image, is extracted and transformed to 63 x 63 RDM, 4) this process is reiterated 10 times and the best agreement across the 10 RDMs produced at each repetition is computed. These steps are repeated for each number of estimated topics, thus producing a 3-Dimensional 63 pictures x 63 pictures x 100 topics RDMs. Note here that we increased the number of topics estimated to ensure that we were not missing some forms of local maximum (given that a higher number of topics might capture specific lexical fields related to History and War for instance).

References

- Bartlett, F. (1932). *Remembering : a study in experimental and social psychology*. Cambridge New York: Cambridge University Press.
- Halbwachs, M. (1992). *On collective memory*. Chicago: University of Chicago Press.
- Hirst, W., Yamashiro, J. K., & Coman, A. (2018). Collective Memory from a Psychological Perspective. *Trends in Cognitive Sciences*, 22(5), 438–451. <http://doi.org/10.1016/j.tics.2018.02.010>
- Coman, A., Brown, A. D., Koppel, J., & Hirst, W. (2009). Collective Memory from a Psychological Perspective. *International Journal of Politics Culture, and Society*. <http://doi.org/10.1007/s10767-009-9057-9>
- Barnier, A. J., & Sutton, J. (2008). From individual to collective memory: Theoretical and empirical perspectives. *Memory*, 16(3), 177–182. <http://doi.org/10.1080/09541440701828274>
- Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, 113(29), 8171–8176. <http://doi.org/10.1073/pnas.1525569113>
- Gagnepain, P., Vallée, T., Heiden, S., Decorde, M., Gauvain, J.-L., Laurent, A., ... Eustache, F. (forthcoming). Collective memory shapes the organization of individual memories in the medial prefrontal cortex.
- Assmann, J., & Czaplicka, J. (1995). Collective Memory and Cultural Identity. *New German Critique*, (65), 125. <http://doi.org/10.2307/488538>
- Olick, J. K. (1999). Collective Memory: The Two Cultures. *Sociological Theory*, 17(3), 333–348.
- Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, 21(8), 618–631. <http://doi.org/10.1016/j.tics.2017.04.013>
- van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, 35(4), 211–219. <http://doi.org/10.1016/j.tics.2012.03.001>

[//doi.org/10.1016/j.tins.2012.02.001](http://doi.org/10.1016/j.tins.2012.02.001)

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114. <http://doi.org/10.1016/j.neuropsychologia.2013.11.010>

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision Making. *Neuron*, *76*(6), 1057–1070. <http://doi.org/10.1016/j.neuron.2012.12.002>

Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*(3), 103–109. <http://doi.org/10.1016/j.tics.2008.12.005>

Wagner, D. D., Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The Dorsal Medial Prefrontal Cortex Responds Preferentially to Social Interactions during Natural Viewing. *Journal of Neuroscience*, *36*(26), 6917–6925. <http://doi.org/10.1523/jneurosci.4220-15.2016>

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*(4), 1694–1703. <http://doi.org/10.1016/j.neuroimage.2004.04.015>

Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual Review of Psychology*, *62*(1), 103–134. <http://doi.org/10.1146/annurev-psych-120709-145406>

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, *3*, 993–1022. Retrieved from <http://dl.acm.org/citation.cfm?id=944919.944937>

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation.. *Psychological Review*, *114*(2), 211–244. <http://doi.org/10.1037/0033-295x.114.2.211>

Steyvers, M., & Griffiths, T. L. (2007). Probabilistic topic models. In *Handbook of Latent Semantic Analysis*. Hillsdale, NJ: Erlbaum.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. <http://doi.org/10.1371/journal.pcbi.1003553>

Pereira, F., Detre, G., & Botvinick, M. (2011). Generating Text from Functional Brain Im-

- ages. *Frontiers in Human Neuroscience*, 5. <http://doi.org/10.3389/fnhum.2011.00072>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458. <http://doi.org/10.1038/nature17637>
- Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Frontiers in Psychology*, 3. <http://doi.org/10.3389/fpsyg.2012.00245>
- Abdi, H., Williams, L. J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 124–167. <http://doi.org/10.1002/wics.198>
- Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2013). Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology*, 3. <http://doi.org/10.3389/fpsyg.2012.00606>
- Pennebaker, J. W., & Gonzales, A. L. (2009). Making History: Social and Psychological Processes Underlying Collective Memory. In P. Boyer & J. V. Wertsch (Eds.), *Memory in Mind and Culture* (pp. 171–193). Cambridge University Press. <http://doi.org/10.1017/cbo9780511626999.010>
- Overwalle, F. V. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858. <http://doi.org/10.1002/hbm.20547>
- Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4), 451–470. <http://doi.org/10.1002/wcs.1183>
- Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2016). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20(1), 115–125. <http://doi.org/10.1038/nn.4450>
- van Kesteren, M. T. R., Fernandez, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, 107(16), 7550–7555. <http://doi.org/10.1073/pnas.0914892107>

- Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the Neural Correlates of Schemas: Ventromedial Prefrontal Cortex Is Necessary for Normal Schematic Influence on Memory. *The Journal of Neuroscience*, *35*(47), 15746–15751. <http://doi.org/10.1523/jneurosci.2767-15.2015>
- van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359. <http://doi.org/10.1016/j.neuropsychologia.2013.05.027>
- Halbwachs, M. (1925/1994). *Les cadres sociaux de la memoire*. Paris: Albin Michel.
- Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*. <http://doi.org/10.3389/fnsys.2013.00074>
- Hirst, W., & Rajaram, S. (2014). Toward a social turn in memory: An introduction to a special issue on social memory. *Journal of Applied Research in Memory and Cognition*, *3*(4), 239–243. <http://doi.org/10.1016/j.jarmac.2014.10.001>
- Wagoner, B. (2013). Bartlett's concept of schema in reconstruction. *Theory & Psychology*, *23*(5), 553–575. <http://doi.org/10.1177/0959354313500166>
- Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. <http://doi.org/10.3389/neuro.06.004.2008>
- Turner, B. O., Mumford, J. A., Poldrack, R. A., & Ashby, F. G. (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*, *62*(3), 1429–1438. <http://doi.org/10.1016/j.neuroimage.2012.05.057>
- Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. <http://doi.org/10.1016/j.neuroimage.2011.08.076>
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. <http://doi.org/10.1016/j.neuroimage.2015.12.012>
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix,

N., ... Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1), 273–289. <http://doi.org/10.1006/nimg.2001.0978>

Peschanski, D. (2012). Régimes de mémorialité et conditions de la mise en récit mémoriel. In *Les Années noires*. Hermann, Paris.

Heiden, S. (2010). The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In *24th Pacific Asia Conference on Language, Information and Computation* (pp. 389–398). Sendai, Japan: Institute for Digital Enhancement of Cognitive Development, Waseda University. Retrieved from <https://halshs.archives-ouvertes.fr/halshs-00549764>

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Supplement 1), 5228–5235. <http://doi.org/10.1073/pnas.0307752101>

” *"Mieux vaut voir une chose toujours pour la première fois que la connaître, Parce que connaître c'est comme n'avoir jamais vu pour la première fois, Et n'avoir jamais vu pour la première fois c'est ne savoir que par oui-dire."*

— **Fernando Pessoa**
(Le gardeur de troupeaux)

Nous avons présenté ici trois études qui ont abordé l'influence de l'inhibition et des schémas collectifs sur la mémoire humaine. Nous allons discuter ici, au-delà des implications directes de chacun de ces travaux qui sont rappelées dans leur discussions respectives, des perspectives sur lesquelles ils peuvent déboucher. Nous mentionnerons en particulier leur apport quant à la compréhension de l'oubli pris en un sens assez large, et notamment sur la nature et la réalité de la suppression des souvenirs et des émotions qui peuvent y être associées. Nous évoquerons également l'importance des résultats issus de la troisième étude dans l'intégration des travaux neuroscientifiques à l'étude de la mémoire collective. Nous nous intéresserons en particulier non seulement aux modèles neurobiologiques qui permettent d'étayer ces résultats mais aux perspectives, déjà, qu'ils peuvent ouvrir sur l'étude de l'oubli, pris cette fois en un sens collectif, qui fera également écho aux réflexions précédentes concernant la suppression volontaire de certains souvenirs.

5.1 La permanence du souvenir

Les études 1 et 2 de ce travail se sont particulièrement concentrées sur l'étude de l'oubli volontaire chez des populations de sujets jeunes ne présentant pas de troubles psychiatriques identifiés. Le paradigme "*Think/No-think*" qui nous a servi de cadre expérimental tout au long de ces travaux aborde la question de l'oubli par le biais de l'inhibition volontaire du rappel des souvenirs ciblés, ce qui en fait une forme d'oubli bien identifiée et qui se distingue d'autres formes telles que l'oubli dû au déclin ou à la sélection des traces mnésiques que nous avons détaillé en introduction. Toutefois, bien qu'il existe plusieurs modalités qui aboutissent à une forme d'oubli, la question de la permanence des traces mnésiques sous une forme altérée, pouvant potentiellement réapparaître à l'avenir, est un commentaire fréquent pour de telles recherches, et même redoutable si l'on considère l'ampleur des questions qu'il soulève lorsqu'on tente d'y répondre de façon définitive.

Ces remarques pourraient presque se résumer à ce que propose l'aphorisme 126 d'Aurore, écrit par Nietzsche en 1881 : "*Il n'est pas encore démontré que l'oubli existe ; tout ce que nous savons c'est qu'il n'est pas en notre pouvoir de nous ressouvenir*" (Nietzsche, [1881/2012](#)). Cette critique n'est en réalité jamais complètement

écartée lorsqu'il est question d'étudier l'oubli, et elle peut toujours se voir réactiver en réponse aux études cherchant à mettre en évidence une disparition complète de celui-ci, en suggérant que la trace de l'expérience, ou même la trace de cette trace de l'expérience reste sous une forme silencieuse, encore présente. Mais il faut aussi voir que, sur le plan des neurosciences, cette argumentaire ne va pas sans une surenchère constante dans la définition d'une trace mnésique sur laquelle nous pouvons nous accorder. Pouvoir argumenter sur la disparition "complète" d'une trace mnésique revient en effet à tenter de prouver une absence, ce qui reste difficile s'il n'est pas possible de s'accorder sur une méthode qui permettrait de révéler la présence de celle-ci. La neuropsychologie s'accorde pourtant sur un ensemble d'influences que devraient pouvoir exercer les traces mnésiques, telles que la possibilité d'y accéder consciemment, l'influence comportementale qu'elle peuvent avoir, les réaction émotionnelles qui peuvent y être liées etc. . . Toutes ces manifestations pouvant être mesurées objectivement. Nous pouvons donc considérer par ce biais qu'une trace mnésique "inconsciente", qui ne pourrait plus jamais se manifester ni dans l'accès en mémoire ni dans une quelconque influence des comportements ou des émotions, présenterait toutes les marques d'une suppression complète de la mémoire alors même que la démonstration de cette suppression serait un processus délicat.

Ce soupçon envers l'oubli que mentionnait Nietzsche, qui influencera aussi Freud et la psychanalyse qui suivra, a aussi été avancé à de nombreuses reprises dans les neurosciences cognitives (voir (Anderson et al., 2011) pour une discussion de ces points concernant le paradigme "Think/No-Think"), et doit aussi être considéré pour les implications qu'il peut avoir sur le plan neurobiologique. On peut notamment se demander si les souvenirs qui sont supprimés lors d'un oubli induit par la suppression du rappel sont oubliés en raison d'une difficulté d'accès ou d'un effacement de la trace. La difficulté à accéder à un ensemble de souvenirs n'implique pas nécessairement une suppression de la trace. L'utilisation de l'imagerie fonctionnelle a jusqu'à présent apporté des arguments forts en faveur de la suppression, notamment en montrant l'impact de cet oubli sur les zones corticales encodant les détails épisodiques de ces souvenirs (Gagnepain et al., 2014). Pourtant, le coût supplémentaire que peut impliquer l'effacement du souvenir par rapport à l'impossibilité de son accès pourrait être un argument en faveur d'une implication de l'accès plutôt que de la suppression complète des traces. L'économie du cerveau est à la fois contrainte par un besoin de conservation de l'énergie, ce qui implique que la réorganisation complète d'un réseau de neurone devrait être évitée si elle n'est pas nécessaire, et par la nécessité d'une maximisation de l'utilisation des neurones qui le composent et donc d'une réutilisation des constituants des traces mnésiques dont il se débarrasse.

Cette dernière hypothèse semble toutefois être contredite par quelques études récentes ayant suggéré une conservation latente de traces mnésiques alors même qu'elles semblaient avoir été supprimées sur le plan neurobiologique. Récemment, une étude (Guskjolen et al., 2018) a rapporté la réactivation de souvenirs encodés durant l'enfance chez des souris par manipulation optogénétique. L'amnésie infantile est, chez de nombreuses espèces, due au moins en partie à une forte activité de

neurogenèse dans l'hippocampe, ce qui en fait un modèle de suppression mnésique d'origine physiologique particulièrement intéressant. De nombreux mécanismes peuvent intervenir sur le plan biologique pour expliquer les différentes formes d'oubli que nous avons mentionnées (Davis and Zhong, 2017).

Les circuits neuronaux qui sont engagés dans la suppression de souvenirs qui est observée lors du “Think/No-Think” restent encore mal compris. Ils semblent toutefois impliquer le cortex cingulaire antérieur, qui est une composante importante de la régulation “top-down” exercée sur le lobe temporal médian et notamment l'hippocampe (Anderson et al., 2016). La concentration en acide γ -aminobutyrique (GABA) dans l'hippocampe est notamment liée à la capacité des individus à supprimer volontairement les souvenirs (Schmitz et al., 2017), ce qui suggère que les inter-neurones gabaergiques pourraient être les relais dans cette zone de la régulation de l'activité et entraîner par extension la suppression des souvenir. Le concept de “trace mnésique” peut véhiculer un certain nombre de confusions lorsqu'on tente d'en aborder l'implémentation au sein de réseaux de neurones. En particulier, il se pourrait que ces traces ne se présentent pas d'un seul tenant mais impliquent chacune deux réseaux interconnectés se maintenant réciproquement en équilibre. On considère que le système nerveux utilise des ensembles de cellules excitatrices pour encoder les souvenirs d'expériences passées, mais un ensemble d'études suggèrent aussi que des “engrammes inhibiteurs” sont générés en réponse à la formation de ces réseaux pour en assurer la stabilité et prévenir la réactivation chaotiques de ces souvenir (Barron et al., 2017). Ceux-ci pourraient par exemple être impliqués dans la réduction de la réponse comportementale et/ou neurale face à des stimuli familiers. A l'inverse, une désinhibition pourrait sous-tendre les mécanismes du rappel de ces souvenir. Les neurones gabaergiques sont aujourd'hui considérés comme des constituant probables de ces “engrammes inhibiteurs”. La modulation du ratio excitation/inhibition peut, par exemple, être influencée par une stimulation trans-crânienne à courant direct. Et il a récemment été montré que cette procédure peut induire la ré-expression d'associations mnésiques qui avaient atteint une forme latente suite à un phénomène d'habituation (Barron et al., 2016). L'étude de l'implémentation neurobiologique de la suppression volontaire pourrait donc à l'avenir tenter d'explorer l'influence du contrôle inhibiteur sur la réactivation spontanée des souvenir. La modulation “top-down” de l'hippocampe par le cortex préfrontal droit pourrait ainsi induire localement, soit une sur-inhibition de ces “engrammes inhibiteurs”, soit une complète désinhibition qui pourrait rendre le réseau instable. Les conséquences sur l'activité du réseau excitateur, qui est censé encoder l'influence de l'expérience passée, pourraient être très différentes selon ces deux alternatives.

5.2 Une régulation émotionnelle des souvenirs

Si nous avons rapporté un oubli induit par l'inhibition du rappel dans nos deux études, tout du moins une incapacité à rappeler l'image concernée, la première étude s'est également penchée sur la régulation physiologique des émotions qui

étaient associées à ces souvenirs. La réaction physiologique associée à l'émotion de dégoût était modifiée lors de la dernière présentation pour les images ayant été volontairement écartées, dont le rappel a été répétitivement inhibé lors de la tâche de "Think/No-Think". Un ajout important a été apporté par la partie de l'étude portant sur le second groupe qui, en plus de répliquer les résultats précédemment obtenus, montrait que cet effet ne pouvait pas être expliqué par des fluctuations lentes de la fréquence cardiaque, induites par exemple par la présentation groupée des images d'une même condition, mais était une réponse directe induite par la présentation d'une image appartenant à la condition de suppression volontaire du rappel. C'est en tout cas ce que suggéraient les données issues d'une présentation randomisée des items. Ce résultat implique aussi que la modulation physiologique, que nous avons interprétée comme positive dans le sens d'un rapprochement avec une condition neutre, est une réponse à la perception du stimulus, et par conséquent une reconnaissance de celui-ci, ce qui pourrait indiquer que le souvenir n'a en réalité pas disparu.

Nous avons discuté dans l'article n°1 les aspects méthodologiques et les liens avec la *théorie polyvagale* (Porges, 1995) et le *modèle d'intégration neuroviscérale* (Thayer and Lane, 2000), et tenté de justifier la validité de cette observation. Pourtant, l'origine même de cette modulation et le lien qu'elle entretient avec la capacité de contrôle des intrusions ouvrent sur un vaste ensemble de questions. D'une part en ce qui concerne son origine : ce contrôle de la fréquence cardiaque est-il réellement lié à l'oubli des souvenirs ? La régulation émotionnelle peut à la fois être réalisée à travers une suppression complète de la réaction initiale, ou par la mobilisation conjointe d'une réaction antagoniste qui pourrait masquer cette dernière. Comme nous l'avons mentionné, le paradigme "Think/No-Think" que nous avons implémenté se prête mal à la comparaison de réactions liées aux items oubliés contre les items non oubliés dû à un faible nombre d'oublis d'une manière générale, et à l'absence complète d'oubli pour certains participants. Nous avons à l'inverse cherché à comparer les sujets ayant beaucoup oublié les items de la condition intrusive, que cet oubli soit lié à la suppression ou non, et montré qu'une plus grande capacité de contrôle de la variabilité de la fréquence cardiaque était liée à un plus grand oubli de ces items. Toutefois, cette mesure ne permet pas encore d'inférer que les items les plus oubliés sont ceux qui ont induit une plus grande modulation de la fréquence cardiaque. Lors de l'analyse du second groupe intégré à cette même étude, la diminution de la fréquence Thêta pour les items de la condition No-Think comparés à la condition Think nous a permis de répliquer un résultat central rapporté par d'autres études (Waldhauser et al., 2014; Ketz et al., 2014), en montrant que cette diminution est directement liée à la modulation moyenne de cette variabilité de la fréquence cardiaque.

Ces éléments permettent de lier l'inhibition du rappel et la modulation subséquente de la réponse physiologique, mais n'indiquent pas directement que la suppression des souvenirs est la cause de la régulation émotionnelle. Une hypothèse alternative pourrait avancer que cette modulation du système nerveux

autonome est induite par la répétition du besoin d'inhibition du rappel qui aurait induit un apprentissage et une automatisation d'une réaction de contrôle. Le modèle d'intégration neuroviscérale, dans sa formulation la plus récente (Smith et al., 2017), propose par exemple une description computationnelle des interactions entre les différentes hiérarchies de contrôle entre le système nerveux central et le système nerveux périphérique (voir Figure 5.1). Plus spécifiquement, ce modèle propose d'appliquer la perspective du cerveau bayésien, qui stipule que le cerveau est un système complexe spécialisé dans la représentation probabiliste de niveaux de certitudes et d'incertitudes concernant les sensations, l'attention et le contrôle de l'organisme (Knill and Pouget, 2004). Ce cadre théorique intègre notamment un ensemble d'hypothèses concernant des architectures cérébrales impliquées dans le codage prédictif (Friston, 2010). Dans ce cadre théorique, les processus simples comme la reconnaissance d'un stimulus n'engagent pas des mécanismes passifs de détection de certaines caractéristiques à travers un ensemble de "filtres", mais la production et la mise à l'épreuve de modèles internes. La perception et l'attention sont alors alimentées par la concordance entre les prédictions de ces modèles et ce que les sensations peuvent fournir. Il est également proposé que ces prédictions opèrent selon différents niveaux des hiérarchies corticales pour tenter de prédire les signaux reçus par les structures inférieures. Ces modèles sont susceptibles de générer de hauts niveaux d'erreurs si leurs prédictions s'avèrent erronées, et peuvent les adapter en continu de façon à fournir un modèle du monde, ou des ensembles de signaux qu'ils sont amenés à traiter, la plus fiable possible. Smith et al., 2017 ont alors proposé que les différentes hiérarchies du système d'intégration neuroviscérale puissent être impliquées dans ce type de prédictions.

Dans le cadre de notre étude, nous considérons que l'origine de la régulation cardiaque que nous avons rapportée se trouve dans les niveaux les plus élevés de cette hiérarchie, dans le réseau de contrôle exécutif. Celui-ci pourrait potentiellement interagir avec le réseau par défaut ou l'amygdale dans l'adaptation de la réponse émotionnelle. Cela pourrait par exemple consister en une reconfiguration des représentations a priori (*priors*) associées aux images négatives de sorte que leur présentation à la conscience, qu'elle soit induite par une intrusion mnésique ou par une perception directe, est associée à une réaction automatisée de contrôle. L'implémentation de cette réaction, si elle est avérée, et les circuits qu'elle emprunte, sont encore difficiles à définir. L'augmentation relative de la fréquence cardiaque pour les images issues de la condition No-Think comparée aux images issues de la condition contrôle pourrait suggérer une activation du système nerveux sympathique ou une désactivation du système nerveux parasympathique. Compte-tenu des dynamiques temporelles qui sont rapportées (<5s), nous avons considéré que le système sympathique ne pourrait être mobilisé dans un intervalle aussi court et qu'une réduction brusque de l'activité parasympathique pourrait davantage expliquer cette modulation.

Il est à noter que si cette réaction semble effectivement liée à l'application d'un contrôle inhibiteur répété, qu'elle est concomitante à l'oubli des images par la suite

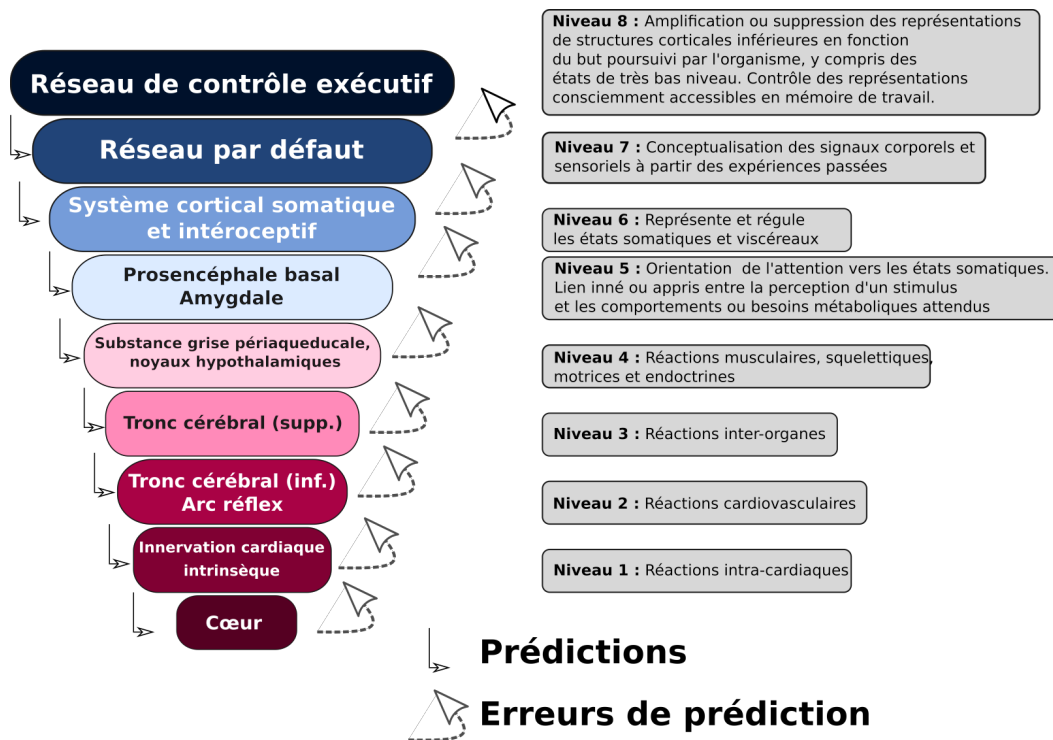


Fig. 5.1: Modèle d'intégration neuroviscérale. Le modèle d'intégration neuroviscérale propose que la régulation de la fréquence cardiaque peut être opérationnalisée à différents niveaux d'une hiérarchie comprenant 8 niveaux dans sa version la plus récente, partant de mécanismes entièrement autonomes propres à l'innervation du cœur jusqu'aux structures frontales les plus récentes dans l'évolution, engagées dans les comportements orientés et l'accès à la conscience des représentations. Chaque structure est engagée dans une minimisation des erreurs de prédiction des inputs des structures inférieures. Adapté de Smith et al., 2017.

et, dans une moindre mesure, à un effet d'amélioration de l'évaluation subjective des images rapporté dans l'étude n°2, la régulation physiologique semble impliquer l'encodage d'une réaction automatisée plutôt que la suppression directe d'une activité particulière. C'est en effet l'un des paradoxes du contrôle de la mémoire en vue de générer un contrôle des émotions : la suppression effective du souvenir pourrait potentiellement ramener l'organisme à un stade de pré-exposition à l'événement et supprimer tout phénomène d'habituation qui pourrait avoir un effet comportemental bénéfique. Supprimer la représentation négative pourrait ainsi induire une plus grande réaction émotionnelle lorsque cet élément sera à nouveau perçu.

5.3 Le rythme des intrusions

Nous avons rapporté dans la première étude une diminution généralisée de l'activité Thêta (3-8 Hz) dans les conditions de contrôle volontaire des souvenirs par rapport la condition de rappel volontaire. En nous basant sur le rôle documenté de cette fréquence dans l'encodage et la récupération de souvenirs épisodiques (Backus et al., 2016; Fell and Axmacher, 2011; Kerrén et al., 2018), ainsi que sur le rapport

de résultats similaires dans un contexte expérimental très proche de celui que nous avons utilisé (Waldhauser et al., 2014), nous avons interprété cette réduction comme le reflet d'une réduction de la réactivation des traces que le participant devait inhiber. Ces résultats pourraient sembler contraster avec les analyses réalisées dans la seconde étude qui ont montré l'implication de l'activité Thêta enregistrée au niveau des électrodes préfrontales dans le contrôle des intrusions. Cette observation fait également suite à de nombreuses études qui ont rapporté le rôle de cette fréquence dans le contrôle cognitif (Eschmann et al., 2018; Cavanagh and Frank, 2014), notamment en réaction à la détection d'interférences dans le rappel de souvenirs épisodiques (Staudigl et al., 2010; Hanslmayr et al., 2010). Il importe premièrement de noter que les différences qui sont rapportées ici résultent de la comparaison de conditions différentes : le contraste entre une condition de rappel volontaire et une condition d'inhibition du rappel volontaire d'une part (première étude), et le contraste entre une condition d'inhibition du rappel en réaction à une intrusion prédite par notre modèle et cette même condition sans intrusion (seconde étude). Il n'est donc pas incohérent de considérer que l'inhibition du rappel volontaire est d'une façon générale liée à une réduction de l'activité Thêta, comparé au rappel qui nécessite possiblement un contrôle cognitif important. Mais d'autre part, les intrusions mnésiques survenant alors même que le participant tente de les prévenir peuvent être sous-tendues par cette même activité, qui peut alors être engagée dans leur contrôle.

Une des difficultés principales qui peut se présenter pour l'interprétation de ces résultats se trouve dans la conception oscillante ou constante à la fois des intrusions mnésiques et du contrôle cognitif. Il pourrait être avancé que les participants qui n'ont rapporté aucune intrusions sur certains essais l'ont fait grâce à un haut niveau de contrôle cognitif exercé durant toute la durée de celui-ci (3 secondes). Ces deux hypothèses (oscillant vs. maintenu), appliquées aux intrusions et aux mécanismes de contrôle, peuvent produire au moins quatre modèles que nous tentons de résumer dans la Figure 5.3. Il peut parfois être avancé que la réactivation de souvenirs est un mécanisme continu, apparaissant selon une logique de tout ou rien, et il serait également tentant de dire que l'oubli répond aux mêmes dynamiques. Pourtant, un grand nombre d'études ont documenté, tant chez l'Homme que chez l'animal, que la recombinaison des détails épisodiques de souvenirs, telle qu'elle apparaît lors du rappel, fluctue selon des oscillations dont les périodes sont très en dessous de la seconde. En particulier, la fréquence Thêta est apparue, dans l'hippocampe, être une activité essentielle pour la synchronisation de réseaux distants et l'alternance de phases propices à l'encodage et de phases propices à la récupération (Kerrén et al., 2018; Hasselmo et al., 2002; Nyhus and Curran, 2010). Nous interprétons les résultats de la seconde étude en faveur d'un modèle supposant à la fois une intrusion et un contrôle inhibiteur apparaissant de façon ponctuelle. Ce modèle est cohérent avec d'autres résultats montrant un processus de détection et de résolution des intrusions rapide (environ 500 ms) (Hanslmayr et al., 2010).

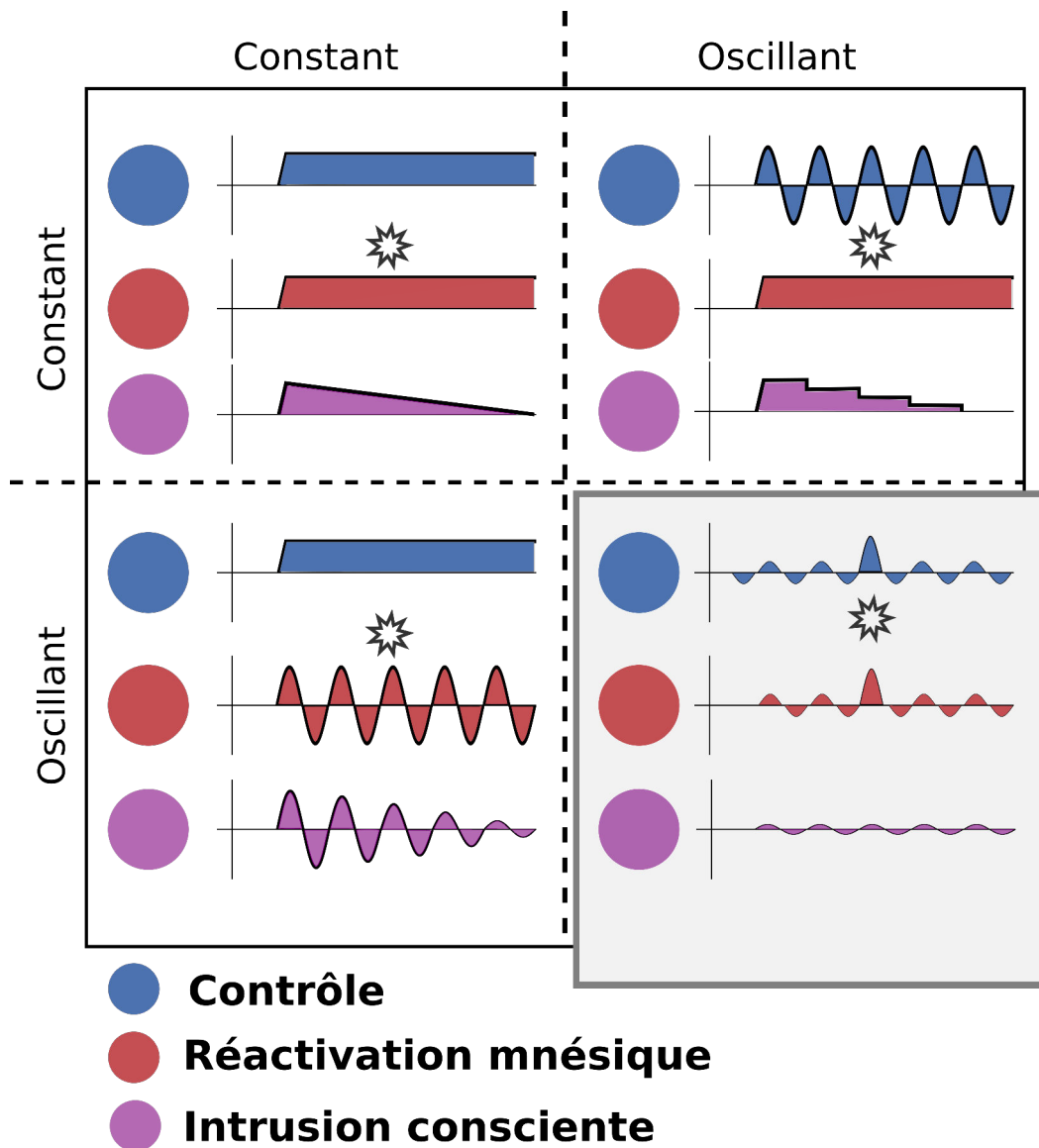


Fig. 5.2: Dynamiques possibles du contrôle inhibiteur et des intrusions mnésiques dans la régulation des intrusions. Selon que l'on considère la réactivation des traces mnésiques et le contrôle inhibiteur comme résultant de processus oscillants ou constants, leur influence sur le traitement des intrusions peut être différente. Nous représentons ici le contrôle inhibiteur en bleu et la réactivation involontaire de souvenirs en rouge. Les courbes violettes représentent l'interaction entre ces dynamiques produisant la perception consciente des intrusions. Les résultats présentés dans la seconde étude sont cohérents avec l'idée d'une manifestation oscillante à la fois de la réactivation mnésique et du contrôle inhibiteur.

L'ordre dans lequel ces mécanismes sont supposés intervenir ne peut toutefois pas être déterminé à partir de nos résultats. Nous nous sommes ici essentiellement concentrés sur l'activité Thêta. Cette fréquence, qui est une des plus bases du spectre d'intérêt de l'analyse en électroencéphalographie, est également susceptible d'interagir avec d'autres fréquences plus élevées à travers un phénomène de couplage inter-fréquence (Jirsa and Müller, 2013). Ceci qui peut impliquer

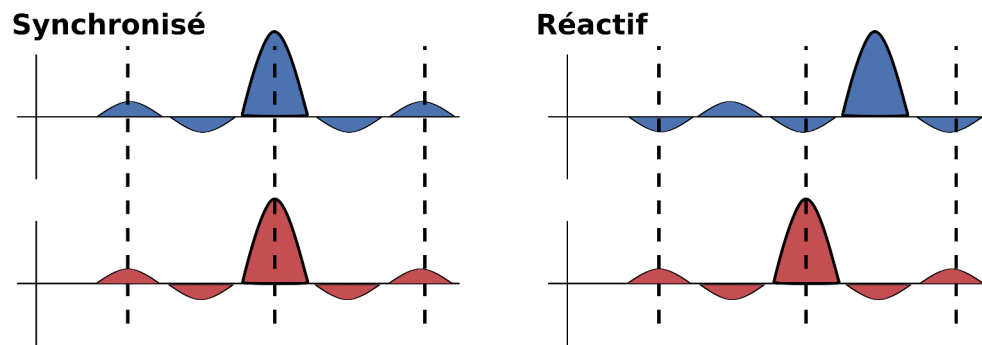


Fig. 5.3: Dynamiques possibles du contrôle inhibiteur et des intrusions mnésiques dans la régulation des intrusions. Ces manifestations oscillantes peuvent apparaître de façon synchronisées l'une par rapport à l'autre ou en réaction et de façon séquentielle. Les résultats que nous avons présentés ne permettent pas de favoriser l'une ou l'autre de ces alternatives.

que l'amplitude ou la phase d'une fréquence "base" va influencer l'amplitude ou la phase de fréquence plus "élevées". Ce mécanisme est également omniprésent dans la synchronisation de réseaux neuronaux et le transfert d'informations entre des zones distantes. De nombreux travaux ont par exemple rapporté le rôle du couplage entre la phase de la fréquence Thêta et l'amplitude des fréquences Gamma dans la reconstitution des détails épisodiques (Heusser et al., 2016; Lega et al., 2014) ou l'activité des cellules de lieu (Jirsa and Müller, 2013). De façon intéressante, le contrôle inhibiteur pourrait être réalisé par des mécanismes de couplage entre fréquences ou par la réinitialisation de phases de certaines fréquences au niveau du cortex préfrontal (Helfrich and Knight, 2016). Nous pourrions donc avancer l'hypothèse selon laquelle l'augmentation ponctuelle de l'activité Thêta frontale que nous avons rapportée sous-tend une réaction intrinsèque de haute fréquence qui pourrait permettre une désynchronisation volontaire de ces réseaux en intervenant lors de phases propices à la réactivation mnésique. La nature de ces oscillations reste, dans le cas de tâche comme le "Think/No-Think", encore inconnue. Nous pouvons simplement mentionner l'importance des activités Bêta et Gamma qui a été soulignée par des études récentes pour la mémoire de travail (Lundqvist et al., 2016; Lundqvist et al., 2018) ainsi que dans le contrôle inhibiteur d'origine préfrontal (Wagner et al., 2018), et l'intérêt que pourrait constituer leur interaction avec les mécanismes que nous avons décrits dans le cadre de futures étude.

5.4 Oublier l'histoire ?

Dans les relations qu'elles peuvent entretenir avec l'oubli et la capacité de contrôle des intrusions mnésiques que nous avons évoqués, les émotions se sont largement rapportées au domaine de l'individuel, pour ne pas dire de l'intime, et ce de façon assez singulière relativement à leur histoire. Les émotions seraient en effet le propre d'un sujet, si bien qu'elles s'y ancrent viscéralement, à travers l'évolution, et qu'elles

semblent avoir existé pour l'Homme de toute éternité. Le terme même est pourtant apparu assez tardivement, au XVII^{ème} siècle, alors que ses proches "émeute" ou "émouvoir" (mettre en mouvement) étaient déjà présents en français. Mais le terme désignait alors une "agitation populaire", et ce n'est qu'au cours du XIX^{ème} siècle que le sens "trouble de l'économie animal" devient premier (Rimé, 2009). Les résultats préliminaires que nous avons présentés dans la troisième étude ont participé à montrer que de la même façon les souvenirs ont une origine et sont intégrés à un cadre collectif qui est par certains aspects premier dans leur encodage. Nous avons abordé la question de l'encodage et de la récupération de représentations historiques qui ont une place importante dans les ressources culturelles auxquelles les participants ont pu être exposés. Les résultats préliminaires de ce travail tendent à confirmer les observations précédentes (Gagnepain et al., forthcoming), réalisées auprès de participants ayant visité le Mémorial de Caen, quand au rôle de la partie dorsale du cortex préfrontal médian dans l'encodage et la récupération de ces représentations.

Ces résultats appuient également l'hypothèse d'une structuration des représentations collectives autour de schémas mnésiques dominants, ainsi que les nombreuses implications qui accompagnent ce modèle. Il est à noter ici que la dimension historique ajoute à ces concepts une temporalité déterminante, mais qui peut permettre également de renouveler certaines approches propres aux sciences humaines et sociales. Nous avons évoqué dans l'introduction de ce troisième travail les trois dimensions qu'il pouvait être données de la mémoire collective, en détaillant pour chacune d'entre elles les implications quant à la modalité d'inscription des traces en mémoire, et au mouvement qui les faisait devenir collectives. Les schémas mnésiques sont des structures cognitives qui peuvent être influencées par ces trois dimensions, encodant la répétition d'événements ou l'exposition à des ressources culturelles, l'influence des mécanismes de partage social des souvenirs et des émotions et les logiques sous-tendant l'émergence de comportements collectifs. Ces différentes formes que peut prendre le souvenir collectif pourraient pourtant être complétées par ce que les neurosciences nous ont appris des schémas, notamment en ce qui concerne leur rôle dans la compréhension des informations présentes dans l'environnement. En réalité, ce "collectif" que nous avons décrit jusqu'à présent à travers ces trois dynamiques a eu une position partielle dans la sélection de son étymologie. Le latin "colligo", dont ce terme est dérivé, contient en quelque sorte deux racines. Il peut s'agir de la concaténation du préfixe "con-" (avec, tous ensemble, tout) et du verbe "ligo" : lier. C'est d'une certaine façon ce sens que nous avons privilégié en montrant que cette mémoire collective est un lien : entre des souvenirs semblables chez différents individus, entre des ressources culturelles et une société d'individus, ou encore que ce lien entre les individus, à travers le réseau, peut former une nouvelle manifestation d'une forme de mémoire globale. Mais il y a aussi un autre verbe qui peut être utilisé pour former ce mot : lego : « lire, cueillir ». Une dualité similaire entre le lien et la lecture existe avec le terme de "religion", c'est la lecture commune des textes qui fait le lien de la religion. De la même manière,

un souvenir a la propriété d'être collectif s'il s'agit d'une représentation que l'on "cueille" ensemble, et nous retrouvons ici l'idée de tri qui intervient dans la sélection des souvenirs lors de leur consolidation, et l'effet du partage à travers la discussion des événements récents, ce sont aussi des souvenirs que l'on "lit" ensemble, dont on a une interprétation et une compréhension qui est partagée, et enfin ce sont des souvenirs sur lesquels on se "recueille" ensemble.

Cette dernière notion trouve un écho important avec le concept de commémoration, et la capacité de pardon et d'oubli dont parle Paul Ricoeur à la fin de son ouvrage *La mémoire, l'histoire, l'oubli* (Ricoeur, 2003). Mais il renvoie aussi à cette idée de retour sur le souvenir, ce phénix de l'événement dont parle aussi François Dosse (Dosse, 2010), et qui évoque le besoin pour mettre en perspective le présent de l'ancrer dans un passé, de le réduire à un déjà-vu. L'histoire doit revenir pour éclairer le sens des événements collectifs, et il est impossible d'extraire le souvenir du réseau mnésique qui lui donne sens. C'est le lien que celui-ci entretient avec d'autres représentations qui va le placer dans un référentiel, mais aussi lui conférer une utilité comportementale. Un souvenir utile est un souvenir qui permet de comprendre la marche des événements auxquels l'organisme est soumis. Ramené aux théories du cerveau bayésien que nous avons évoquées précédemment, un souvenir utile est un souvenir qui permet de minimiser les erreurs de prédictions des événements futurs. Le concept de schéma mnésique ne dénote pas seulement une logique passive d'abstraction d'événements semblables, il rend aussi la mémoire collective tributaire d'un l'ancrage historique face aux événements et au contexte représentationnel qu'ils incarnent. C'est en comprenant que la mémoire collective est aussi une grille de lecture des événements vécus collectivement que nous pouvons aborder sous un nouvel angle l'idée d'un oubli collectif.

L'oubli joue un rôle central dans la construction de la mémoire collective (Brown et al., 2012; Stone and Hirst, 2014). C'est par l'oubli progressif de souvenirs particuliers, incohérents ou peu signifiants que les structures plus larges et partagées des événements apparaissent. Cette forme d'oubli, cette sélection des souvenirs, n'est pas encore un rôle de la capacité volontaire à oublier, cette résistance aux souvenirs que nous avons décrits dans les deux premières parties. Il est possible de faire l'hypothèse qu'il existe un biais social envers certains événements collectifs qui pourraient en favoriser l'oubli tels qu'il est étudié à travers le paradigme "Think/No-Think", de la même manière que certaines émotions peuvent interagir avec cette capacité de suppression. Au-delà de la logique de cohérence des schémas entre eux, et de la conservation des événements structurant l'ensemble des autres souvenirs, la mémoire collective est aussi guidée par sa puissance de fondation d'une identité collective, et à la préservation de l'intégrité morale de cette identité. Des informations qui pourraient remettre en question cette dimension morale du groupe peuvent ainsi être traitées différemment. Il n'existe à l'heure actuelle aucune donnée permettant de conclure que des mécanismes de contrôle inhibiteurs semblables à ce qui est requis pour la suppression d'intrusions mnésiques soient engagés dans de tels contextes, ces questions restant encore largement inexplorées par la psychologie

et les neurosciences sociales. Par ailleurs, l'utilité d'un oubli volontaire en contexte social ne se présente sans doute pas de façon aussi évidente que pour ce qui concerne la régulation de processus cognitifs et émotionnels pouvant être impliqués dans le développement de troubles psychiatriques. Une étude (Coman et al., 2014) a par exemple montré que leurs participants oubliaient plus volontiers les justifications ou les informations contextuelles d'actes moralement condamnables si elles concernent des actions commises par le membre d'un autre groupe, tandis que ces informations ne seront pas oubliées si elles concernent un des membres du groupe. Intégrer une information en contradiction avec les valeurs morales du groupe ne peut se faire sans une adaptation, un ajustement du système à même de les faire cohabiter sans remettre en cause la structure dans son ensemble.

Pourtant, nous pouvons aussi voir qu'une autre forme d'oubli pourrait être impactée par l'évolution induite par l'histoire sur l'organisation des souvenirs au sein même de la mémoire collective. C'est celui par exemple d'une isolation progressive de l'événement des structures qui en ont permis l'intégration. C'est ce qui pourrait se produire lorsqu'il est difficile pour un souvenir particulier d'intégrer le récit mémoriel, c'est-à-dire d'entretenir avec d'autres représentations des liens associatifs suffisamment forts pour qu'elle puissent s'appeler mutuellement sans contradictions et sans représenter une remise en cause de l'intégrité morale de l'individu et de son groupe. Il apparaît alors que la conservation de l'événement dans le temps n'est pas uniquement fonction de son encodage initial mais est aussi influencée par la nature des événements qui l'ont suivi. Ces derniers ont eux-même joué un rôle de reconstruction des schémas associés et potentiellement un effet d'isolation si l'événement ne réapparaît pas comme semblable à d'autres par la suite (voir Figure 5.4). Raymond Aron avait cette formule célèbre : *"Il faut rendre au passé l'incertitude de l'avenir"*. L'inscription en mémoire collective est le produit d'un événement avec ceux qui lui ont succédé. Et à l'inverse, la ré-évoation du contexte d'un souvenir peut avoir un effet de distorsion (Doss et al., 2018). Ceci est particulièrement déterminant lorsque l'on voit en l'histoire des cycles qui pourraient se répéter, que des périodes différentes puissent par certains aspects se ressembler et qu'il soit possible, par un jeu de similarités, de tirer des leçons sur le présent à partir de ces ressemblances. Comme nous l'avons montré, ces ressemblances sont au centre du concept même de schéma : c'est parce qu'il y a ressemblance que l'encodage du souvenir est accéléré par la réduction de ses points communs. Mais cette réduction peut aussi être une certaine forme de violence, à la fois dans l'amoindrissement de la complexité de ce qui est perçu dans le présent, mais aussi par la réévaluation que cela entraîne des représentations pré-existantes. Récemment, une étude a montré que les analogies historiques peuvent être des moteurs importants dans la plasticité de la signification des événements historiques eux-mêmes, allant jusqu'à en ré-évaluer les implications ou à en altérer la représentation (Lee et al., 2017). Si les pages sombres de l'histoire peuvent être aujourd'hui utilisées comme arguments rhétoriques, invectives ou menaces, cela n'a pas simplement pour effet de distordre la perception de ce qui se trouve dans le présent, mais cela peut aussi produire une adaptation insidieuse du

passé pour le présent, et conséquemment une isolation de ce qui, dans l'événement comparé, n'est pas comparable.

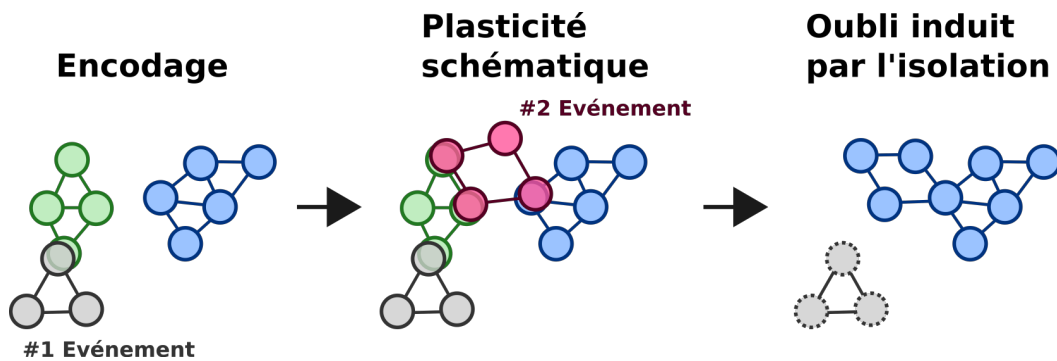


Fig. 5.4: Oubli induit par la reconfiguration des schémas mnésiques. Les représentations des événements historiques passés sont activement utilisées pour interpréter les événements présents et simuler un futur probable. La conservation du souvenir d'un événement historique donné dépendra, en plus de son encodage initial, de la conservation de la structure associative sur laquelle il s'est construit. Cette structure peut être altérée par une succession d'autres événements qui seront davantage liés à certaines représentations, ou à la reconfiguration de ces schémas au travers de leur utilisation dans la simulation d'un futur probable. L'exemple des bombardements alliés sur les villes normandes lors de la libération est une illustration d'une telle isolation progressive installée d'un événement du reste d'un réseau explicatif puissant régulièrement mobilisé. Mais de la même manière, la projection collective vers un futur altéré par le réchauffement climatique peut également induire l'émergence de nouveaux attracteurs qui peuvent occulter ou déstabiliser les réseaux signifiants préexistants.

Ces utilisations du passé pour étayer les considérations morales des comportements collectifs peuvent avoir une influence importante sur la conservation même de ces souvenirs. La mémoire collective joue un rôle central dans les raisonnements impliquant des structures sociales, l'identité collective ainsi que pour notre capacité à donner du sens aux événements récents. L'exposition aux médias est un vecteur important de la réactivation de certains pans de l'histoire face aux actualités, et les représentations présentes en mémoire collective sont des articulations importantes de l'adhésion aux "fake-news" (Bavel and Pereira, 2018). Les recherches en psychologie ont beaucoup insisté sur cette dialectique entre le passé et le présent, en cherchant finalement assez peu à expliquer la fonction prospective de la mémoire collective (Saint-Laurent and Obradović, 2018). La neuropsychologie a depuis une dizaine d'année exploré ces relations entre la mémoire rétrospective, la mémoire des expériences passées et la capacité à se projeter mentalement dans un avenir probable (Schacter et al., 2007; Schacter and Addis, 2007; Viard et al., 2011). Pour la même raison que l'on parle alors de pensée du futur ou de mémoire du futur, on peut projeter la mémoire collective dans le futur que se représentent ses individus (Szpunar and Szpunar, 2016). Récemment, les recherches en mémoire collective se sont penchées sur la relation entre l'influence de ces représentations du passées et la capacité à se projeter sur le plan collectif, par exemple dans un futur

altéré par le changement climatique (Shrikanth et al., 2018; Merck et al., 2016). On comprend que de telles projections ne vont pas sans une sélection, potentiellement très influente, de certains souvenirs ou récits qui permettraient d'en alimenter les dynamiques. Ainsi, la protection des mémoires collectives contre l'oubli, est aussi un contrôle, une régulation de la reconstruction constante des souvenirs autour de l'actualité; et il n'y a guère de menaces plus imminentes pour la conservation des mémoires que la promesse d'affronter un futur qui ne leur ressemble plus.

Rêves et mémoire : revue de la littérature en neuropsychologie et neurosciences cognitives

Dreams and memory: review in neuropsychology and cognitive neurosciences

Nicolas Legrand^{1,2,3,4},
Fabian Guénolé^{1,2,3,4,5},
Francis Eustache^{1,2,3,4},
Géraldine Rauchs^{1,2,3,4}

¹ Inserm, U1077, GIP Cycleron, boulevard Becquerel, BP 5229, 14074 Caen cedex 5, France

² Université de Caen Normandie, UMR-S1077, 14000 Caen, France

³ École pratique des hautes études, UMR-S1077, 14000 Caen, France

⁴ CHU de Caen, UMR-S1077, 14000 Caen, France
<geraldine.rauchs@inserm.fr>

⁵ CHU de Caen, Service de psychiatrie de l'enfant et de l'adolescent, 14000 Caen, France

Pour citer cet article : Legrand N, Guénolé F, Eustache F, Rauchs G. Rêves et mémoire : revue de la littérature en neuropsychologie et neurosciences cognitives. *Rev Neuropsychol* 2015 ; 7 (4) : 279-90 doi:10.1684/nrp.2015.0362

Résumé

Le rêve fait l'objet d'une grande fascination pour la psychologie et les neurosciences depuis le XIX^e siècle. Pourtant, en dépit d'avancées récentes, il reste encore profondément énigmatique. Décrit comme un état modifié de conscience survenant au cours du sommeil, caractérisé par des expériences sensorielles, cognitives et émotionnelles vives, il est largement admis que les rapports de rêves sont une traduction des souvenirs de l'expérience rêvée. Toutefois, de nombreux obstacles méthodologiques et théoriques freinent encore les neuroscientifiques dans leurs recherches sur les mécanismes de production et de mémorisation de ces expériences. Ces dernières années, de nombreuses études ont décrit les effets bénéfiques du sommeil sur la mémoire, réactivant l'idée selon laquelle les processus de consolidation des souvenirs pourraient alimenter l'activité onirique. Pourtant, de façon surprenante, les recherches montrent aussi que les rapports de rêves ne peuvent que rarement être reliés à des souvenirs épisodiques précis, et que leur contenu ou leur fréquence ne semblent pas clairement associés aux performances mnésiques. Nous proposons une revue des recherches sur les liens entre rêves et mémoire en abordant deux questions : (1) Peut-on déterminer quels types de souvenirs réapparaissent au sein des rêves ? (2) Les rêves jouent-ils un rôle dans la consolidation et la réorganisation de la mémoire lors du sommeil ? Nous détaillons ensuite un modèle décrivant l'assimilation, lors du sommeil, des souvenirs récents au sein de structures mnésiques globales en cohérence avec les données issues de l'étude neuroscientifique du rêve et les connaissances fournies par l'analyse des récits. Nous soulignons enfin l'importance des rapports fournis par les rêveurs pour l'étude des processus de consolidation mnésique et leur apport à la neuropsychologie dans l'étude du vieillissement et des pathologies neurodégénératives.

Mots clés : rêve • mémoire • amnésie • consolidation • vieillissement

Abstract

Dreaming is the subject of a huge fascination in psychology and neuroscience since the 19th century. However, despite recent advances, this phenomenon remains deeply enigmatic. Described as a modified state of consciousness occurring during sleep, characterized by vivid sensory, cognitive and emotional experiences, it is widely accepted that dream reports are the reflection of memories of dream experiences. However, numerous methodological and theoretical pitfalls are still slowing neuroscientists in their researches on production and memorization of such experiences. During the last decade, numerous studies have described the beneficial effect of sleep on memory, bringing back the idea that memory consolidation processes could underpin dreaming. However and surprisingly, studies

Correspondance :
G. Rauchs

reveal that dream reports can barely be related to specific episodic memories, and neither content nor frequency seems clearly related to memory performance.

Here, we review studies that focused on the relationship between dreams and memory, and address two main questions: (1) Can we define the type of memories that are replayed during dreams? (2) Do dreams play a role in consolidation and reorganization of memory during sleep? We detail a model describing the assimilation of recent memories in global memory frameworks during sleep, in line with neuroscientific data from dream studies and knowledge from dreams reports. We finally emphasize the importance of dreamer's reports for the understanding of memory consolidation processes, and their contribution to neuropsychology in normal aging and neurodegenerative diseases.

Key words: dreams • memory • amnesia • consolidation • aging

■ Introduction

La signification et le rôle des rêves ont été au centre d'un intérêt constant pour de nombreuses disciplines comme la philosophie dès l'Antiquité, la psychologie expérimentale ainsi que la psychanalyse à partir du XIX^e siècle, mais aussi plus récemment pour la neuropsychologie et les neurosciences [1]. Pourtant, les connaissances dans ce domaine restent encore très limitées et sont l'occasion de nombreuses controverses théoriques qui attribuent aux rêves des fonctions psychologiques parfois très différentes. Cela s'explique par le fait que le rêve se prête peu à une approche scientifique du fait de sa difficulté d'accès. Après qu'un sujet en a fait l'expérience (*rêve-vécu*), le rêve doit être encodé en mémoire pendant le sommeil et rappelé au réveil (*rêve-souvenir*), ce souvenir souvent fugace peut alors être retranscrit en mots ou en images (*rêve-récit*). Cependant, seule cette troisième étape du processus de rappel est réellement observable et mesurable dans le cadre d'études scientifiques [2]. La mémoire occupe donc une place importante dans ces expériences, à la fois lors de l'encodage du souvenir du rêve, dans la production de celui-ci lors du sommeil en permettant l'accès à certaines représentations, mais également aujourd'hui comme hypothèse quant à la fonction même de l'activité onirique. Certaines études ont en effet proposé que les rêves pourraient refléter certains mécanismes de consolidation mnésique lors du sommeil [3-5]. Cependant, les liens entre la capacité à rappeler un rêve au réveil et les performances en mémoire ne semblent pas clairement établis. Nous ferons dans un premier temps la revue des travaux ayant étudié les liens entre les troubles de la mémoire et l'absence de rappel de rêves. Nous aborderons ensuite l'étude de la réapparition des souvenirs dans les rêves en nous concentrant sur l'hypothèse de leur consolidation lors de ces expériences. Nous présenterons un modèle de réorganisation des traces mnésiques lors du sommeil basé sur les résultats d'études récentes, et soulignerons enfin l'intérêt de l'étude des rapports de rêve dans la neuropsychologie du vieillissement normal et pathologique.

■ Amnésie et anonirie

En neuropsychologie, les liens entre les rêves et la mémoire ont tout d'abord été abordés à travers leurs altérations respectives [6]. L'anonirie désigne une cessation complète des rapports de rêves qui se traduit par une incapacité de l'individu à rappeler de telles expériences, même s'il est réveillé lors d'une phase de sommeil paradoxal. Ce trouble est l'un des symptômes figurant dans le syndrome de Charcot-Willebrand, décrit en 1880, où l'on observe notamment une atteinte de la jonction temporo-pariétale et du cortex préfrontal médian. L'étude de 112 patients présentant ce trouble a montré que celui-ci peut être entièrement dissocié des performances mnésiques telles qu'elles sont évaluées par les tests neuropsychologiques classiques [6]. Mais l'une des difficultés majeures lors de tels cas est de déterminer dans quelle mesure cette absence de rappel est due à un problème de mémorisation des rêves ou à une réelle absence de production de ceux-ci lors du sommeil.

L'identification de zones cérébrales connues pour jouer un rôle dans la mémoire épisodique lors de la production de ces expériences est également difficile à déterminer. L'un des premiers modèles neurobiologiques suggérait que les rêves n'apparaissent que lors du sommeil paradoxal (*REM sleep*) en étant générés par des décharges neuronales chaotiques provenant des régions pontiques [7], ce qui faisait du contenu onirique le résultat d'un processus imprévisible relativement indépendant des souvenirs du sujet. Pourtant, d'autres travaux ont montré que le rêve ne se limite pas à cette phase de sommeil. Ainsi, si un sujet réveillé en sommeil paradoxal rapporte un rêve environ huit fois sur dix, un réveil en sommeil lent donne lieu à un rappel dans la moitié des cas [8]. Ce modèle a été plus récemment remis en cause par l'étude des lésions ayant produit soit une cessation du sommeil paradoxal, soit une absence de rappel de rêves [9]. Ces résultats ont montré une dissociation fonctionnelle entre l'activité propre au sommeil paradoxal, qui dépend des zones pontiques, et la capacité à rappeler des rêves, qui dépend de régions corticales (*i.e.* la jonction temporo-pariétale et le cortex préfrontal médian) [10].

La jonction temporo-pariétale et le cortex préfrontal médian sont impliqués dans de nombreuses fonctions cognitives et sont communs à plusieurs réseaux fonctionnels cérébraux. Des études en imagerie ont suggéré que les régions qui sous-tendent l'activité onirique pourraient comprendre une sous-partie du réseau par défaut (*Default Mode Network* [11]). Ce réseau est largement impliqué dans la récupération de souvenirs épisodiques, la projection dans le futur ou encore la théorie de l'esprit [12]. L'hippocampe fait partie de ce réseau et se trouve activé avec le cortex entorhinal et les régions parahippocampiques lors du sommeil paradoxal [13]. Du fait de l'importance de cette zone dans l'encodage, la consolidation et la récupération de souvenirs épisodiques, il a été suggéré qu'elle pourrait sous-tendre la réactivation de certains souvenirs lors de la production de rêves [14]. Pour autant, l'anosmie ne semble pas être engendrée par des altérations de la mémoire dues à des lésions, que ce soit dans les parties postérieures du cerveau [15], au sein des lobes temporaux [16], ou touchant spécifiquement l'hippocampe [17]. De manière générale, il n'existe pas de corrélation entre des troubles de la mémoire à long terme, qu'ils soient rétrogrades ou antérogrades, et la fréquence de rappel de rêves [6]. Une grande fréquence de rappel ne dénote pas non plus de meilleures performances en mémoire déclarative : les sujets narcoleptiques, qui rapportent énormément de rêves, ont également des plaintes en mémoire prospective et rétrospective [18].

Les études ayant exploré l'effet de lésions spécifiques ont principalement manipulé des variables reflétant la fréquence de rappel de rêves. Or, il apparaît que la dimension qualitative de l'activité onirique (nature et origine des souvenirs, charge émotionnelle, bizarrerie, complexité, longueur, vivacité de l'expérience...) pourrait être plus affectée par ces lésions cérébrales que la fréquence du souvenir de rêve. L'une des rares études portant sur des patients présentant des lésions hippocampiques bilatérales [17] a montré que, si les rapports de rêves perdurent, ils sont néanmoins plus courts, simples, répétitifs et stéréotypés et présentent une dimension émotionnelle moins importante par rapport aux récits de sujets contrôles. La présence d'émotions fortes, qui caractérise la majorité des récits de rêves, serait liée à une augmentation de l'activité de l'amygdale lors du sommeil paradoxal [13] qui provoquerait l'activation des aires para- limbiques. Une étude a par exemple rapporté que l'intégrité microstructurale de l'amygdale gauche est liée à la longueur des rêves rapportés par les sujets et à leur charge émotionnelle [19]. L'activation de ces structures pourrait ainsi, par le biais de l'hippocampe, induire la récupération de certains souvenirs émotionnels particuliers. Le fait qu'une telle dimension qualitative soit absente lors de lésions hippocampiques [17] ou de certains cas d'amnésie [20] montrerait que l'accès aux souvenirs est une condition pour que les rêves présentent des caractéristiques phénoménologiques et psychologiques classiques.

Par ailleurs, une amnésie ne sera pas suffisante pour empêcher tout rappel de rêve. Si des amnésies rétrogrades ou antérogrades peuvent influencer le contenu des récits,

elles ne provoquent pas une cessation complète du rappel de rêves [9]. De plus, si des régions cérébrales impliquées dans la mémoire épisodique semblent déterminantes pour l'encodage du souvenir de l'expérience [14], un bon fonctionnement de la mémoire ne sera pas nécessairement associé à un plus grand nombre de souvenirs de rêves. Les sujets qui présentent une amnésie antérograde ne se souviennent généralement pas de leurs rêves, mais peuvent néanmoins en rapporter s'ils sont réveillés directement lors d'une phase de sommeil paradoxal. Cette dissociation entre les réseaux impliqués dans l'encodage et le rappel de souvenirs et les bases neurales du rappel de rêve a récemment été étayée par la première étude en imagerie qui a permis de contraster l'activité cérébrale en phase de sommeil paradoxal de sujets présentant une faible ou une grande fréquence de rappel de rêve [10]. La comparaison des deux groupes a montré une plus grande activité lors du sommeil paradoxal et à l'état de veille au niveau de la jonction temporo-pariétale et du cortex préfrontal médian chez les personnes rapportant beaucoup de rêves. La jonction temporo-pariétale est impliquée dans les mécanismes d'orientation attentionnelle, ce qui pourrait suggérer que les grands rapporteurs de rêves seraient plus sensibles aux stimuli extérieurs lors du sommeil et pourraient ainsi avoir plus d'éveils intra-sommeil favorisant l'encodage du souvenir des rêves [21]. Le cortex préfrontal médian est, quant à lui, impliqué dans de nombreuses fonctions comme l'encodage et la récupération en mémoire mais aussi la théorie de l'esprit. Les rêves impliquant souvent des interactions entre personnages, ces capacités pourraient intervenir, particulièrement dans la création des rêves les plus complexes. Il reste cependant difficile de déterminer si ces zones clés sont responsables de la production du rêve, de l'encodage en mémoire de son expérience, ou s'ils favorisent simplement l'apparition de courtes phases d'éveils lors du sommeil qui permettront ensuite l'inscription de l'épisode en mémoire à long terme.

En résumé, les capacités mnésiques des sujets ne semblent pas liées à la fréquence des rêves qui sont rapportés. Si certaines zones impliquées dans la récupération de souvenirs épisodiques (cortex préfrontal médian, jonction temporo-pariétale, hippocampe et amygdale) semblent influencer certaines caractéristiques de l'expérience des rêves, l'atteinte de ces régions n'implique pas une cessation complète des rappels de rêves. Ces observations suggèrent ainsi que, si le rêve ne peut être réduit à une activité propre à la mémoire lors du sommeil, l'apparition de ces expériences met en jeu la récupération de certains souvenirs.

■ Quels sont les souvenirs qui réapparaissent le plus fréquemment dans les rêves ?

Selon certains chercheurs, les souvenirs des sujets influencent la production de rêves et ainsi le contenu de

rappports obtenus au réveil. L'idée selon laquelle le contenu d'un rêve reflète des expériences vécues la veille ou les jours précédents a été avancée dès le XIX^e siècle par Freud et Hervey de Saint-Denys, et a été fortement réactivée à la fin des années 1950, après la description du sommeil paradoxal [22]. De nombreuses études ont par la suite confirmé cette hypothèse grâce à l'étude des sources mnésiques des rêves [14] et en tentant de décrire la nature des liens entre les états mentaux diurnes et le contenu des rêves de la nuit suivante [23].

Les chercheurs ont testé dans quelle mesure il était possible d'influencer le contenu d'un rêve, par exemple en présentant aux sujets des films émotionnellement marquants juste avant l'endormissement [24]. De manière surprenante, ces expositions avaient peu d'influence sur les contenus des rêves rapportés. À l'inverse, si l'on modifiait à l'éveil la perception des sujets, par le biais de lunettes teintées ou d'un prisme inversant le champ visuel [25], ces altérations avaient tendance à persister dans les rêves des nuits suivantes. Ces premières études soulignaient ainsi davantage la permanence d'une altération perceptive diurne dans les rêves ensuite rapportés mais n'impliquaient pas la réapparition de souvenirs clairement identifiables.

D'autres recherches ont ensuite montré que la pratique intensive d'un jeu vidéo comme *Tetris* a une influence directe sur les rêves qui apparaissent lors de l'endormissement suivant la pratique [16]. Cette assimilation de sources très récentes est fréquemment observée dans les rêves de première partie de nuit, dominée par du sommeil lent, et de façon encore plus prononcée lors des hallucinations hypnagogiques qui surviennent à l'endormissement [26, 27]. Cet effet est suffisamment important pour être rapporté même par des sujets amnésiques. Stickgold *et al.* [16] ont par exemple montré que la pratique intensive du jeu *Tetris* peut induire des hallucinations hypnagogiques chez des patients présentant des lésions bilatérales dans le lobe temporal médian, notamment l'hippocampe, alors même que ceux-ci n'étaient pas en mesure de former un souvenir explicite de la pratique. Ces observations indiquent que la formation d'un rêve n'implique pas nécessairement les zones classiques de la mémoire déclarative mais peut se construire à partir de traces sensorielles laissées par une expérience récente, indépendamment du fait que celles-ci donnent lieu ou non à un souvenir épisodique.

Toutefois, si les souvenirs d'épisodes vécus la veille, ce que Freud nommait les « *résidus diurnes* », sont essentiels dans la création du contenu d'un rêve et tendent à être réactivés en début de nuit (*i.e.* en sommeil lent), ces éléments ne suffisent pas pour décrire des rêves complexes survenant de façon plus tardive, notamment en sommeil paradoxal. Si les travaux chez le rat ont montré que les traces associées à un apprentissage sont effectivement réactivées dans les premières phases de sommeil, cette activité tend à décroître au cours de la nuit [28]. Les rêves en sommeil paradoxal sont par ailleurs plus créatifs, ils présentent

une structure narrative élaborée, incluent potentiellement plusieurs personnages et un contenu émotionnel fort [29]. L'identification des souvenirs qui y interviennent est alors plus difficile et peut impliquer des sources non seulement récentes (datant du jour précédent) mais aussi plus anciennes (particulièrement entre 5 et 7 jours) [30]. Cette incorporation d'événements plus anciens a été interprétée comme le reflet de processus de consolidation spécifiques qui ne découleraient pas d'une réactivation automatique des traces mais impliqueraient un mécanisme d'intégration au sein d'aires corticales, modulé par la charge émotionnelle et l'importance du souvenir pour le rêveur [14, 31].

Pourtant, il est encore difficile de déterminer quelle proportion de l'activité onirique représentent les restes diurnes relativement à des éléments plus anciens. Si l'étude de journaux de rêves, souvent tenus pendant de grands intervalles de temps, révèle certaines régularités dans la réactivation des sources mnésiques, il n'est cependant pas possible de prévoir le contenu des rêves sur la simple base des expériences récentes [23]. Nous savons, par exemple, que les sujets dormant au sein d'un laboratoire de sommeil ont tendance à intégrer des références à ces lieux dans 6 à 32 % des rêves qu'ils rapportent la nuit suivante [32]. Mais les éléments de la veille ne sont jamais représentés en intégralité, ce qui fait de la réapparition d'un souvenir épisodique au sens plein un phénomène relativement rare [33]. L'idée même d'une continuité entre rêve et éveil peut être décelée à plusieurs niveaux, selon des liens sémantiques, symboliques ou dans la reproduction de structures narratives. De ce fait, les processus d'interprétation des rapports de rêve qui peuvent être utilisés dans le cadre scientifique sont encore mal définis.

Une étude a tenté de déterminer quelle était la part de l'expérience du rêve qui pourrait être reliée de façon fiable à des souvenirs d'expériences récentes en comparant le contenu de carnets de rêves avec le contenu de carnets de journée (description du contenu de la journée par le participant) [33]. Parmi près de 300 récits de rêves recueillis, il est apparu que seulement 1 à 2 % des scènes de rêves rappelées pouvaient être considérées comme de véritables souvenirs épisodiques. Plusieurs facteurs peuvent être proposés pour expliquer cette faible proportion, notamment le fait que le sommeil paradoxal altère la connectivité entre les zones impliquées dans le rappel et l'encodage d'informations comme l'hippocampe et le cortex préfrontal [5]. Mais Schwartz [34] a aussi montré que les critères utilisés pour l'identification d'un souvenir au sein d'un carnet de rêves sont déterminants. Un souvenir épisodique se caractérise par la reviviscence de la scène avec tous les éléments de contexte associés, tels que le lieu, les personnages, les objets, les émotions, etc. Or, si la réactivation de tous ces éléments ne s'observe que très rarement dans les rêves rapportés en milieu ou en fin de nuit, ceux-ci intègrent très souvent des fragments de souvenirs. Ainsi, 65 % des 300 récits recueillis contenaient certaines composantes du souvenir épisodique du rêveur sans pour autant avoir représenté la scène dans son intégralité.

Une expérience ne réapparaît que très rarement dans un rêve telle qu'elle a été vécue par le sujet. Les rêves les plus complexes qui apparaissent généralement en fin de nuit ont souvent peu à voir avec les expériences issues de la veille. Les activités les plus fréquentes peuvent être sous-représentées, en particulier si celles-ci sont routinières ou demandent une grande concentration (lire un livre, faire ses comptes, etc.), au profit d'activités souvent plus sociales et chargées émotionnellement [35]. Le contenu du rêve est moins une simple réexpression d'expériences passées que le mélange de fragments de souvenirs récents et lointains en fonction d'émotions, des représentations liées à l'identité personnelle ou encore des préoccupations ou des désirs du rêveur [36]. Certains rêves peuvent ainsi évoquer des souvenirs parfois très distants. Grenier *et al.* [37] ont notamment observé que la réapparition de souvenirs dans les rêves obéit à des régularités comparables à celles observées pour la mémoire autobiographique. Les souvenirs les plus récents, mais aussi ceux qui remontent au pic de réminiscence (adolescence et début de l'âge adulte), sont plus fréquemment intégrés dans les rêves. Les seules circonstances dans lesquelles les rapports semblent correspondre précisément à des souvenirs épisodiques sont certaines conditions pathologiques comme l'état de stress post-traumatique [38].

Les recherches actuelles s'accordent à dire que le rêve n'est pas un simple processus de rappel et qu'il n'est pas possible d'y observer la réapparition de souvenirs tels qu'ils ont été vécus. Mais en analysant la réapparition de ces fragments, des études ont observé que, parmi les événements de la vie éveillée, ceux qui datent de sept à huit jours sont plus fréquemment intégrés dans les rêves que ceux qui datent de trois à quatre jours [39]. Ce phénomène, appelé « *dream lag effect* », n'est observable que pour les rêves issus du *sommeil paradoxal* et non ceux issus du *sommeil lent* [30]. De plus, ce délai d'incorporation ne s'applique pas à l'ensemble des souvenirs mais davantage aux événements émotionnellement saillants et importants pour le rêveur [31]. Les raisons de cet aspect de l'expérience onirique restent encore mal comprises. Mais, en considérant que la réactivation d'un souvenir est une étape clé dans le processus de consolidation, l'incorporation différée d'expériences émotionnellement saillantes ou importantes a été interprétée comme le résultat d'une période permettant l'intégration des souvenirs épisodiques dans le néocortex [14]. Cette hypothèse rejoint un ensemble plus large de théories développées ces dernières années en neurosciences et qui postulent que les contenus des rêves sont un reflet des processus de consolidation, de tri et d'intégration des souvenirs en mémoire [3, 40]. Ce cadre théorique, qui se trouve alimenté par les nombreuses études portant sur le rôle du sommeil dans la mémoire, pourrait permettre d'expliquer, à partir de ces mécanismes, les particularités du contenu des rapports de rêve. Ces hypothèses ont ainsi en commun de suggérer que les apprentissages récents et complexes tendent à réapparaître dans les contenus des rêves rapportés par la suite. Mais certaines études avancent même l'idée selon laquelle le rêve serait, non pas le reflet des étapes de

consolidation ou d'intégration en mémoire, mais l'une de ces étapes qui jouerait un rôle fonctionnel dans le renforcement des apprentissages et souvenirs récents.

■ Le rêve permet-il de consolider les souvenirs récents ?

La fonction précise des rêves est un sujet largement débattu. De nombreuses hypothèses ont été proposées, comme une fonction de stimulation de la créativité, une régulation émotionnelle ou encore comme nous le proposons ici, une fonction d'intégration de souvenirs en mémoire à long terme. Si un accès aux représentations en mémoire ne semble être ni nécessaire, ni suffisant pour le rappel de rêves, les fragments de souvenirs épisodiques semblent pourtant être des éléments centraux dans leur formation. Par ailleurs, de nombreuses études ont démontré l'effet bénéfique du sommeil sur la consolidation des souvenirs en mémoire à long terme et en ont précisé les mécanismes [2, 3, 41 pour revues]. Si tous les stades de sommeil sont bénéfiques pour la mémoire, le sommeil à ondes lentes bénéficie particulièrement à la mémoire épisodique. En revanche, le sommeil paradoxal favoriserait davantage la consolidation des souvenirs émotionnels et les apprentissages procéduraux. Les recherches dans ce domaine ont conduit à reconsidérer les processus oniriques et leur utilité pour la mémoire. Certaines études ont avancé que le rêve, étant le contenu de la pensée correspondant à tout ou partie de l'activité cérébrale pendant le sommeil, refléterait la réactivation et la réorganisation associées à la consolidation de la mémoire [3, 40]. Des modèles récents ont proposé que le rêve révélerait la création, lors du sommeil, d'associations entre des fragments de souvenirs épisodiques sous l'influence de composantes émotionnelles [3].

Cette hypothèse s'appuie notamment sur le fait qu'au plan neurobiologique les traces mnésiques correspondant aux informations nouvellement acquises sont réactivées lors du sommeil lent, et que cette réactivation est associée chez l'homme à une amélioration des performances lors de tests au réveil [43]. Sur le plan psychologique, l'expérience de l'apprentissage peut être ré-exprimée de manière fragmentaire dans le contenu du rêve, sans que l'on sache toutefois si cette réexpression est systématique [4, 26, 44]. Une étude récente a par exemple montré que certains fragments d'expériences peuvent être réactivés en fonction de l'environnement du rêveur. La présentation d'odeurs à des sujets lors du sommeil paradoxal peut induire la réapparition en rêve d'images qui étaient auparavant associées aux odeurs [45]. Il est important de noter ici que l'ensemble des informations acquises lors de la veille ne sont pas systématiquement intégrées en mémoire, et que des mécanismes de tri et d'oubli interviennent en plus des mécanismes de consolidation lors du sommeil [46]. Certains auteurs ont d'ailleurs proposé que le rêve ne reflète pas la répétition d'expériences en vue d'une consolida-

tion, mais au contraire la suppression d'éléments devenus inutiles ou indésirables [47]. Cette idée est renforcée par l'observation que les souvenirs des expériences oniriques sont très labiles et tendent à disparaître rapidement s'ils ne sont pas reconsolidés. S'il reste très difficile de démontrer expérimentalement que la réexpression fragmentaire d'un souvenir dans le rêve est associée à une consolidation des informations concernées, certaines données tendent à invalider l'hypothèse selon laquelle le sommeil paradoxal pourrait participer à une fonction d'oubli. Une étude récente a par exemple montré, en utilisant un paradigme d'oubli volontaire lors duquel les participants doivent s'efforcer de ne pas récupérer certaines informations en mémoire, que le sommeil paradoxal ne renforce pas les mécanismes d'oubli [48]. Si, à l'état de veille, la mise en place d'une inhibition de la récupération tend à supprimer progressivement les souvenirs concernés, il est apparu que le sommeil paradoxal pouvait avoir un effet compensateur et reconsolider ceux-ci, empêchant ainsi leur oubli.

Si le rêve joue effectivement un rôle dans la consolidation de la mémoire, on pourrait alors s'attendre à ce que les apprentissages récents soient ré-exprimés dans les rêves des nuits suivantes. Nous savons que les récits de rêves collectés après une phase de sommeil lent contiennent davantage de sources de type épisodique où interviennent des fragments de scènes vécues, plus ou moins récentes, avec une faible implication du rêveur. Les rêves obtenus après un réveil en seconde partie de nuit, où le sommeil paradoxal prédomine, contiennent en revanche plus d'émotions, associées à des sources sémantiques et autobiographiques [49]. Il est pourtant assez difficile de déterminer *via* l'étude des souvenirs d'expériences oniriques si la réapparition de certains fragments est simplement due à leur récurrence, favorisant ainsi leur accès, si elle a pour fonction leur consolidation, ou si les souvenirs sont intégrés dans un processus de reconstruction de scènes mentales visant à favoriser l'apprentissage et le développement de comportements. La répétition d'apprentissages moteurs semble, par exemple, être fréquente lors du sommeil paradoxal chez les animaux, comme cela peut s'observer à la suite de la suppression de l'atonie musculaire propre à cette phase [50]. L'hypothèse de la simulation de la menace [51] propose dans ce sens que les rêves auraient pour fonction de préparer les individus à l'affrontement de situations dangereuses en permettant la répétition « *offline* » des conduites adéquates. Dans ces circonstances, les souvenirs ne seraient pas réactivés pour être renforcés, mais pour créer un environnement favorable aux apprentissages et à la créativité. Le rôle de cette répétition onirique de comportements élaborés peut être en partie testé en confrontant les sujets à des tâches complexes avant de courtes siestes. On constate alors que des éléments ont plus de chances d'apparaître dans un rêve s'ils sont associés à un apprentissage intense. De Koninck *et al.* [52] ont notamment observé, en suivant un groupe de quatre jeunes femmes intégrées dans une classe française durant six semaines, que celles qui progressèrent le plus rapidement avaient incorporé dans leurs

rêves des éléments de français avant les autres et rapportaient aussi plus de dialogues. D'autres études ont montré qu'à la suite d'un apprentissage intensif en réalité virtuelle, les sujets peuvent rapporter, lors de siestes après l'épreuve, des rêves reproduisant l'environnement ou la tâche réalisée. Cette incorporation a été observée avec un entraînement à un jeu vidéo [16], une simulation de ski [27] ou encore une tâche de navigation dans un labyrinthe [26]. Toutefois, ces réactivations ont essentiellement été décrites sur des phases de sommeil qui suivent immédiatement l'endormissement (stade 1) et qui ne sont pas connues pour jouer un rôle dans la consolidation de la mémoire. Mais d'autres résultats, provenant de l'étude des journaux de rêves tenus pendant dix jours après avoir participé à une tâche de navigation dans un labyrinthe [53], suggèrent qu'il serait possible de les observer sur des phases de sommeil plus tardives, bien que plus rarement.

Au-delà de la simple réactivation de ces apprentissages, les études actuelles cherchent aussi à évaluer si des sujets qui ont des rêves en lien avec une tâche réalisée auparavant ont de meilleures performances au réveil que les sujets qui n'en ont pas. Wamsley *et al.* [26] ont par exemple observé, grâce à une tâche de navigation dans un labyrinthe virtuel, que les sujets qui ont rapporté un rêve en lien avec la tâche d'apprentissage lors de siestes proposées entre les sessions ($n = 4/50$) ont progressé davantage dans les tests suivants que ceux qui n'ont pas rapporté de tels rêves ($n = 46/50$) ou qui se sont simplement représenté l'environnement de jeu mentalement mais sans dormir ($n = 2/49$).

Ces quelques résultats n'excluent pas l'hypothèse d'un effet de l'utilisation de certaines fonctions cognitives spécifiques sur le contenu des rêves qui sont ensuite rapportés. Il est pourtant encore impossible avec les données existantes d'établir une relation causale entre le contenu d'un rêve et les performances dans une épreuve au réveil. Ces études demandent à être reproduites sur des nuits entières en contrôlant précisément les rappels au sein de chaque phase de sommeil avant de pouvoir conclure à un effet bénéfique des rêves sur la mémorisation d'un apprentissage. Un seul rappel ne permet d'accéder qu'à de courts fragments de l'activité onirique, et les effets hypothétiques des rêves oubliés au cours de la nuit ne sont pas pris en compte dans les études réalisées jusqu'à maintenant. Par ailleurs, si un tel renforcement existe, il semble concerner des comportements relativement complexes et spécialisés qu'il est difficile d'évaluer. Les épreuves neuropsychologiques classiques ne permettent par exemple pas de corréler les rapports de rêves aux résultats obtenus lors de tests de mémoire [6]. L'objectif des futures recherches sera alors de déterminer si les rapports de rêves peuvent effectivement refléter la consolidation mnésique et si l'apparition d'un état de conscience modifié lors du processus de consolidation mnésique en améliore effectivement la qualité ou s'il ne s'agit que d'un processus entièrement indépendant. Une absence de rêve serait alors le symptôme d'une consolidation moins efficace. Cette seconde hypothèse pose la question du rôle de la conscience du rêve, lors de son

expérience et dans le souvenir qui en est formé, et reste encore aujourd'hui difficilement testable scientifiquement. Il est ainsi impossible de déterminer si un rêve mémorisé et un rêve oublié ont les mêmes effets sur les fonctions cognitives au réveil.

Pour résumer, les difficultés précédemment décrites, à la fois dans l'identification des souvenirs au sein de rapports de rêves et dans l'identification du rôle de celui-ci pour la mémoire, pourraient être imputées à une définition trop générale du processus de consolidation qui présuppose que les expériences récentes sont réactivées dans leur intégralité lors du sommeil. Nous allons voir que des études récentes suggèrent à l'inverse que seuls certains fragments des souvenirs sont consolidés en fonction des représentations qui préexistent en mémoire. La prise en compte de ces mécanismes pourrait alors permettre une meilleure interprétation des récits de rêves et inciterait à considérer le rôle de ces expériences relativement au fonctionnement global des systèmes de mémoire.

■ Le rêve et la dynamique de la mémoire épisodique lors du sommeil

Si la production du rêve emprunte indiscutablement aux souvenirs du rêveur [30, 33, 34], cela reste difficilement interprétable dans le cadre des modèles actuels de la consolidation mnésique. Le rêve est un processus créateur dont les liens avec l'imagination ou les mécanismes de voyage mental sont établis, tant phénoménologiquement qu'en termes de corrélats neuraux [54]. Pour certains auteurs, une dimension essentielle de l'activité onirique réside non seulement dans la réactivation et l'assimilation de nouveaux souvenirs, mais également dans la création de nouvelles associations entre différents éléments de la mémoire. Il a par exemple été montré que le sommeil paradoxal, généralement associé à des rapports de rêves plus longs et plus complexes, semble être bénéfique pour certaines fonctions cognitives. Comparés à des sujets réveillés en sommeil lent, les sujets réveillés dans cette phase de sommeil présentent, en effet, une plus grande créativité dans la résolution de problèmes qui nécessitent l'exploration de solutions originales et la création d'associations nouvelles [55]. Cela pourrait indiquer que les bénéfices induits par le sommeil ne sont pas liés seulement à une consolidation des souvenirs, mais comprennent également une augmentation de la flexibilité et de la créativité *via* une modification des réseaux cérébraux.

La consolidation en mémoire épisodique opère principalement en sommeil lent profond par le biais d'interactions entre l'hippocampe et les aires néocorticales [42]. Cependant, le sommeil n'induit pas seulement une consolidation par la réactivation des traces (renforcement synaptique), mais permet également l'intégration des souvenirs récents aux souvenirs plus anciens et la création de nouvelles associations entre représentations [46]. Ces processus associatifs ont une place centrale dans certains modèles récents

qui suggèrent que les co-activations de souvenirs lors du sommeil se feraient, par exemple, en fonction de liens émotionnels et contextuels entre des fragments de souvenirs [3]. Ces réorganisations des souvenirs lors du sommeil sont déterminantes pour la dynamique de la mémoire puisqu'elles permettent le tri, l'assimilation et l'association de nouvelles informations au sein du système mnésique.

Le concept de *schémas* mnésiques constitue un point de départ intéressant pour décrire les mécanismes qui pourraient intervenir dans cette intégration des nouvelles représentations à travers les différentes phases de sommeil. Les schémas sont des structures cognitives héritées de l'expérience qui intègrent, au sein de réseaux associatifs, les éléments récurrents, essentiels et caractéristiques de certaines situations [56]. Si une expérience se présente de façon répétée, comme par exemple une fête d'anniversaire, l'encodage en mémoire procèdera par la création de *schémas* qui associeront entre eux des éléments qui réapparaissent de façon conjointe, comme par exemple la présence d'invités, d'un gâteau et de bougies. Ces schémas offrent un avantage comportemental puisqu'ils augmentent les capacités de prédiction et d'interprétation de l'environnement en favorisant les liens entre des éléments survenant dans des contextes identiques. Ils facilitent également la mémorisation de nouvelles expériences grâce au pré-encodage des éléments récurrents. L'élaboration des schémas serait en grande partie sous-tendue par le cortex préfrontal ventromédian [56] qui joue un rôle central dans l'intégration des nouvelles informations, et est également activé lors du sommeil paradoxal [57]. En outre, des modèles récents ont suggéré que le sommeil pourrait être impliqué dans la création, la modification et la déconstruction des schémas mnésiques pour permettre l'adaptation et l'intégration des nouveaux souvenirs [58].

Lorsque deux événements qui présentent des similitudes sont encodés en mémoire, les traces corticales en partie imbriquées sont associées dans un premier temps, par le biais de l'hippocampe. Au cours du sommeil lent, celles-ci sont réactivées par l'intermédiaire de ces connexions hippocampo-corticales (*figure 1A*). Cette phase de sommeil permettrait de faire émerger les traces les plus importantes pour l'individu, notamment à travers l'activation de réseaux impliqués dans le circuit de la récompense [59] ou dans la régulation émotionnelle [60], qui pourraient agir comme processus de sélection automatique des représentations réactivées lors des rêves [3, 29]. Si ces réactivations de souvenirs impliquent des réseaux partiellement communs, ceux-ci seront reliés les uns aux autres [58]. Ces phases de réactivation alterneraient aussi avec des phases de dégradation synaptique qui réduisent les connexions les plus faibles entre les traces neurales [61], ne laissant que les réseaux fortement associés, ce qui participe à faire émerger les schémas en mémoire (*figure 1B*). La présence de ce schéma pourra, par la suite, favoriser l'intégration de nouveaux souvenirs si ceux-ci présentent certaines similitudes avec les connaissances présentes (*figure 1C*). L'intégration immédiate de nouvelles informations pourrait en effet être

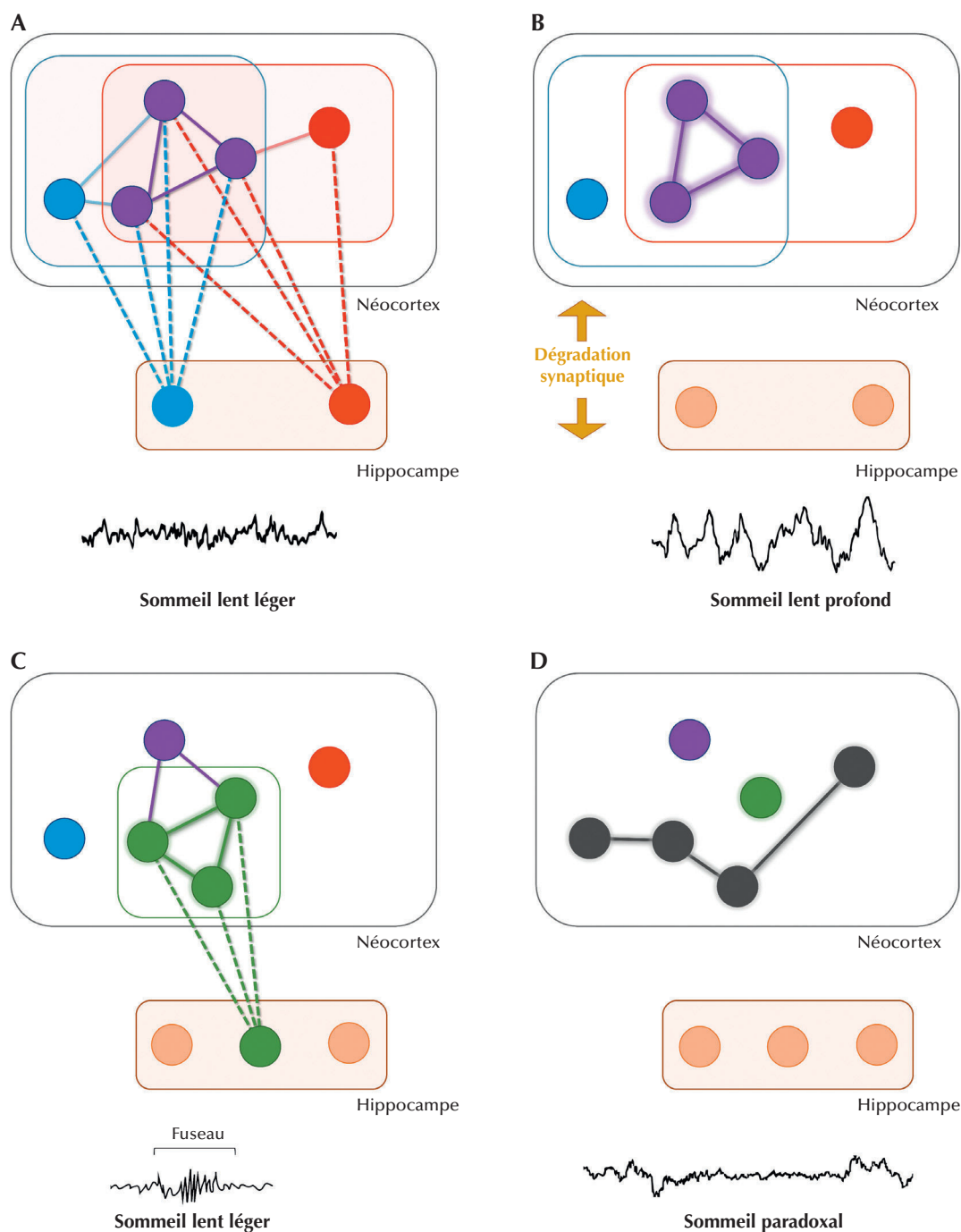


Figure 1. Création et adaptation des schémas lors du sommeil. Chaque sphère représente un groupe de neurones activés lors d'un apprentissage. Lors du sommeil lent, les traces d'expériences récentes sont réactivées par le biais de connexions entre l'hippocampe et le néocortex. Les représentations partagées par plusieurs expériences sont renforcées, ce qui provoque l'apparition de schémas mnésiques (partie A, en violet). Le sommeil lent profond pourrait participer à la dégradation des connexions les plus faibles entre les représentations corticales et ne conserver que les schémas les plus importants (partie B, en violet). Les schémas préexistants pourraient être modifiés par l'intégration de nouvelles informations en mémoire lors de phases de sommeil lent léger, ce qui est reflété par l'apparition de « fuseaux » sur les tracés électroencéphalographiques (partie C, en vert). Le sommeil paradoxal serait un état plus favorable pour la déconstruction et l'adaptation de schémas dans de plus grandes proportions, sous l'influence des émotions et de composantes motivationnelles propres au rêveur (partie D, en noir).

dommageable pour les représentations préexistantes en créant de nombreuses interférences et en répliquant des informations similaires. Tamminen *et al.* [62] ont ainsi montré en utilisant une tâche d'apprentissage de nouveaux mots, que l'activité électroencéphalographique propre au sommeil lent (< 4 Hz) ainsi que les fuseaux de sommeil (éléments caractéristiques du stade 2) sont davantage présents lors de l'intégration de nouveaux concepts si ceux-ci n'entrent pas en interférence avec des concepts déjà appris, en comparaison à l'apprentissage de concepts trop proches des connaissances du sujet. Cette signature électroencéphalographique serait caractéristique des processus de consolidation mnésique intervenant lorsque les éléments à mémoriser présentent suffisamment de cohérence avec les représentations préexistantes.

La réorganisation des représentations peut impliquer davantage qu'une actualisation des schémas en fonction des souvenirs récents et engager la création d'associations nouvelles. Ce mécanisme paraît déterminant lorsque des problèmes complexes doivent être résolus ou lorsque les éléments intégrés paraissent incohérents. Si la facilitation des processus créatifs après le sommeil paradoxal a déjà été soulignée [55], de nombreux exemples, notamment dans l'histoire des sciences, montrent que les rêves pourraient être des reflets relativement précis de ces processus. Des scientifiques, comme le chimiste allemand Friedrich August Kekulé ou le chimiste russe Dmitri Mendeleïev, ont ainsi rapporté des rêves directement liés à certaines de leurs découvertes peu de temps avant qu'ils les réalisent, respectivement la structure hexagonale du benzène et la table périodique des éléments [63]. L'état physiologique pendant le sommeil paradoxal pourrait alors offrir des conditions nécessaires, d'une part, à la déconstruction des représentations préexistantes et, d'autre part, à une re-contextualisation de celles-ci pour permettre l'établissement de nouvelles connexions (*figure 1D*).

La consolidation des souvenirs n'engage donc pas simplement leur réactivation mais aussi une insertion au sein de réseaux préexistants et également une réorganisation de ceux-ci en tenant compte des nouvelles informations assimilées. Cette intégration peut se faire à travers la création ou la modulation de schémas mnésiques qui conservent les éléments les plus récurrents, et peuvent par la suite influencer l'assimilation de nouvelles représentations. La compréhension des étapes successives qui induisent l'encodage à long terme de fragments de souvenirs lors du sommeil permet ainsi une plus grande cohérence, tant avec les données issues d'études comportementales en imagerie ou en électroencéphalographie qu'avec les observations fournies par l'analyse de récits de rêves obtenus le matin ou lors de réveils après différentes phases de sommeil. Cette conception permet d'expliquer plusieurs caractéristiques phénoménologiques des rapports de rêves, leur incohérence et leur étrangeté, mêlant des perceptions vives et des émotions fortes. Enfin, la conservation de structures cognitives intégrant de nouvelles informations peut expliquer que, en dépit de la forte influence des souvenirs récents sur

la formation des rêves, il soit possible d'observer chez une même personne une certaine continuité sur le long terme des éléments ou des thèmes des rêves rapportés.

■ Le rêve au cours du vieillissement normal et pathologique

La capacité à rapporter des rêves semble se conserver tout au long de la vie mais l'évolution de l'activité onirique et l'influence que peuvent avoir sur celle-ci le vieillissement ou les maladies neurodégénératives sont encore peu étudiées. Le vieillissement s'accompagne d'une diminution sensible de la fréquence et de la complexité des rapports de rêves [64]. Ces changements s'installent en réalité dès le début de l'âge adulte, et s'accroissent jusqu'à environ 50 ou 60 ans. Il importe toutefois, pour comprendre ce facteur, de tenir compte de l'évolution des fonctions cognitives impliquées dans le processus de rappel, notamment les capacités mnésiques ou de planification qui sont déterminantes lorsqu'un participant doit tenir un carnet de rêves. L'évaluation rétrospective, par le biais de questionnaires ou la tenue quotidienne d'agendas, doit donc se faire en cohérence avec l'évolution de ces fonctions. L'étude de rappels immédiats après réveils en laboratoire, méthode présentant moins de biais, a toutefois confirmé une diminution globale de la fréquence des rapports de rêves. Cette évolution peut s'expliquer à la fois par des composantes psychologiques, notamment une réduction de l'intérêt des personnes pour leurs rêves ou des conditions de vie moins favorables à leur apparition (ex. réveils imposés tôt le matin avant la fin des cycles de sommeil paradoxal), mais aussi par des modifications dans la physiologie du sommeil. Avec l'âge, la quantité de sommeil lent profond diminue fortement et s'associe à une réduction, plus faible, de la proportion de sommeil paradoxal. Ces changements s'accompagnent d'altérations de certains rythmes circadiens comme l'heure d'apparition des phases de sommeil ou la variation du niveau de cortisol au réveil, qui est impliqué dans la consolidation mnésique. Certaines hypothèses avancent ainsi que les rapports de rêves pourraient fortement dépendre de cette augmentation qui favorise l'encodage des souvenirs [5]. La nature de l'activité cérébrale lors du sommeil paradoxal change également avec l'âge et pourrait affecter les rapports de rêves [65]. Une étude de Chellappa *et al.* [66] a par exemple porté sur les changements de distribution topographique et de fréquence dans les tracés électroencéphalographiques chez des sujets âgés par rapport aux sujets jeunes lors du sommeil. Les résultats ont montré que le sommeil paradoxal est modifié chez le sujet âgé, qui présentent moins d'activité alpha et bêta au niveau frontal que les jeunes.

À l'âge s'ajoutent également de fortes différences en fonction du sexe des sujets. En analysant la fréquence de rappel de rêves de plus de 28 000 personnes, Nielsen [67] a observé que sur une période allant de 15 à 45 ans, les

femmes rappellent significativement plus de rêves que les hommes. Celles-ci montrent un accès au contenu onirique plus précoce et qui se maintient plus longtemps que chez les hommes chez qui il peut commencer à chuter dès 25 ou 30 ans pour atteindre son point le plus bas avant 50 ans, soit dix ans environ avant les femmes. De nombreux facteurs psychologiques ont été évoqués pour expliquer cette différence, comme un plus grand intérêt pour les rêves et de meilleures performances dans le traitement de souvenirs autobiographiques, émotionnels ou sociaux chez les femmes.

Si les mécanismes expliquant les différences de fréquence et de contenu des rapports de rêves dues à l'âge ou au sexe commencent à être étudiés, très peu d'études à ce jour ont abordé l'activité onirique dans le cadre du vieillissement pathologique. De façon générale, une nette diminution de la fréquence des rapports de rêves est observée dans le cadre de pathologies neurodégénératives. Cela semble cohérent, par exemple, pour la maladie d'Alzheimer qui s'accompagne d'une réduction du sommeil paradoxal, d'altérations du sommeil à ondes lentes, de difficultés à l'encodage et d'atteintes structurales au niveau de la jonction temporo-pariétale, zone essentielle dans la fréquence des rapports de rêves. D'autres pathologies, comme la maladie de Parkinson ou la démence à corps de Lewy, s'accompagnent également de troubles du sommeil paradoxal et de l'activité onirique très spécifiques qui ont été davantage décrits ces 20 dernières années. Les rapports de rêves de patients atteints de la maladie de Parkinson tendent à être très vivides, violents et présentent souvent un contenu négatif [68]. Les hommes, qui sont davantage touchés par cette maladie, rapportent davantage de rêves violents que les femmes, différence déjà présente dans des proportions moindres chez le sujet jeune sain. De plus, environ un tiers des patients souffrant de maladie de Parkinson présentent également des troubles du comportement en sommeil paradoxal. Ces troubles sont la conséquence d'une perte de l'atonie musculaire qui conduit les patients à exécuter des mouvements au cours de leur sommeil. Dans de telles conditions, on peut tester si le contenu du rêve correspond au comportement exécuté en sommeil paradoxal [69]. Cette extériorisation semble intervenir surtout dans le cas de rêves prototypiques qui sont souvent plus violents que dans une maladie de Parkinson sans troubles du sommeil paradoxal, et pour une très grande majorité impliquent le sujet dans

des scènes de défense face à des attaques humaines ou animales [70]. Ces manifestations peuvent précéder de plus de dix ans les manifestations motrices de la maladie de Parkinson [68] et sont l'indicateur, dans deux tiers des cas au moins, d'un stade prodromique de cette maladie.

Conclusion

Le rêve est un phénomène complexe qui, tout comme la cognition à l'état de veille, ne saurait être réduit à un seul mécanisme ou à une seule fonction. Les neurosciences contemporaines ont permis l'apparition de concepts nouveaux dans son étude qui ont redéfini les hypothèses sur sa fonction et sa signification. La nature de ces états modifiés de conscience, la construction spontanée de scénarii complexes ou encore l'invention de personnages et de comportements auxquels le sujet se confronte en rêve sont autant d'énigmes dans notre compréhension de la psyché et du cerveau. Les souvenirs semblent jouer un rôle fondamental, à l'origine même du processus de production des rêves. Les mécanismes qui permettent la sélection, la structuration et la réorganisation des représentations en mémoire au cours du rêve sont de mieux en mieux compris, et certaines structures cognitives, telles que les schémas, sont des concepts clés pour comprendre l'implication des souvenirs dans le contenu des rêves, et le rôle éventuel des mécanismes de consolidation dans leur création. Une approche intégrant l'étude du sommeil et de la mémoire aux recherches sur le rêve est indispensable pour mieux comprendre le phénomène autrement inaccessible du rêve. ■

Remerciements

Nicolas Legrand bénéficie d'une allocation doctorale de l'Observatoire B2V des Mémoires.

Nous remercions Perrine Ruby et Béatrice Desgranges pour leurs commentaires critiques sur une version antérieure du manuscrit.

Liens d'intérêts

les auteurs déclarent ne pas avoir de lien d'intérêt en rapport avec l'article.

Références

1. Ruby P. Experimental research on dreaming: state of the art and neuropsychanalytic perspectives. *Front Psychol* 2011 ; 2 : 286.
2. Guénolé F, Nicolas A. Dreaming is a hypnic state of consciousness: getting rid of the Goblots hypothesis and its modern avatars. *Neurophysiol Clin* 2010 ; 40 : 193-9.
3. Llewellyn S. Such stuff as dreams are made on? Elaborative encoding, the ancient art of memory, and the hippocampus. *Behav Brain Sci* 2013 ; 36 : 589-607.
4. Wamsley EJ, Stickgold RJ. Dreaming and offline memory processing. *Curr Biol* 2010 ; 20 : R1010-3.
5. Payne JD. Memory consolidation, the diurnal rhythm of cortisol, and the nature of dreams. *Int Rev Neurobiol* 2010 ; 92 : 101-34.
6. Solms M. *The neuropsychology of dreams*. Mahwah, NJ : Lawrence Erlbaum Associates, 1997.
7. Hobson JA, McCarley RW. The brain as a dream state generator. *Am J Psychiatry* 1977 ; 134 : 1335-48.
8. Nielsen TA. Ultradian, circadian, and sleep-dependent features of dreaming. In : Kryger MH, Roth T, Dement WC, eds. *Principals and Practice Sleep Medicine*, 5th ed. New York : Elsevier, 2011, p. 576-84.

9. Solms M. Dreaming and REM sleep are controlled by different brain mechanisms. *Behav Brain Sci* 2000; 23 : 843-50.
10. Eichenlaub J-B, Nicolas A, Daltrozzo J, et al. Resting brain activity varies with dream recall frequency between subjects. *Neuropsychopharmacology* 2014; 39 : 1594-602.
11. Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 2008; 1124 : 1-38.
12. Legrand D, Ruby P. What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychol Rev* 2009; 116 : 252-82.
13. Braun A. Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study. *Brain* 1997; 120 : 1173-97.
14. Nielsen TA, Stenstrom P. What are the memory sources of dreaming? *Nature* 2005; 437 : 1286-9.
15. Yu CK-C. Memory loss is not equal to loss of dream experience: a clinicoanatomical study of dreaming in patients with posterior brain lesions. *Neuro-Psychoanalysis* 2006; 8 : 191-8.
16. Stickgold R, Malia A, Maguire D, et al. Replaying the game: hypnagogic images in normals and amnesics. *Science* 2000; 290 : 350-3.
17. Torda C. Dreams of subjects with bilateral hippocampal lesions. *Acta Psychiatr Scand* 1969; 45 : 277-88.
18. Wamsley EJ, Donjacour CEHM, Scammell TE, et al. Delusional confusion of dreaming and reality in narcolepsy. *Sleep* 2014; 37 : 419-22.
19. De Gennaro L, Cipolli C, Cherubini A, et al. Amygdala and hippocampus volumetry and diffusivity in relation to dreaming. *Hum Brain Mapp* 2011; 32 : 1458-70.
20. Torda C. Dreams of subjects with loss of memory for recent events. *Psychophysiology* 1969; 6 : 358-65.
21. Eichenlaub J-B, Bertrand O, Morlet D, et al. Brain reactivity differentiates subjects with high and low dream recall frequencies during both sleep and wakefulness. *Cereb Cortex* 2014; 24 : 1206-15.
22. Aserinsky E, Kleitman N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science* 1953; 118 : 273-4.
23. Schredl M. Continuity between waking and dreaming: a proposal for a mathematical model. *Sleep Hypn* 2003; 5 : 26-39.
24. Cartwright R, Bernick N, Borowitz G, et al. Effect of an erotic movie on the sleep and dreams of young men. *Arch Gen Psychiatry* 1969; 20 : 262-71.
25. De Koninck J, Prévost F, Lortie-Lussier M. Vertical inversion of the visual field and REM sleep mentation. *J Sleep Res* 1996; 5 : 16-20.
26. Wamsley EJ, Tucker MA, Payne JD, et al. Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Curr Biol* 2010; 20 : 850-5.
27. Wamsley EJ, Perry K, Djonlagic I, et al. Cognitive replay of visuo-motor learning at sleep onset: temporal dynamics and relationship to task performance. *Sleep* 2010; 33 : 59-68.
28. Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. *Science* 1994; 265 : 676-9.
29. Malinowski JE, Horton CL. Evidence for the preferential incorporation of emotional waking-life experiences into dreams. *Dreaming* 2014; 24 : 18-31.
30. Blagrove MT, Fouquet NC, Henley-Einion JA, et al. Assessing the dream-lag effect for REM and NREM stage 2 dreams. *PLoS One* 2011; 6 : e26708.
31. van Rijn E, Eichenlaub J-B, Lewis P, et al. The dream-lag effect: selective processing of personally significant events during rapid eye movement sleep, but not during slow wave sleep. *Neurobiol Learn Mem* 2015 : 1-12.
32. Schredl M. Laboratory references in dreams: methodological problem and/or evidence for the continuity hypothesis of dreaming? *Int J Dream Res* 2008; 1 : 3-6.
33. Fosse M, Fosse R, Hobson JA, et al. Dreaming and episodic memory: a functional dissociation? *J Cogn Neurosci* 2003; 15 : 1-9.
34. Schwartz S. Are life episodes replayed during dreaming? *Trends Cogn Sci* 2003; 7 : 325-7.
35. Schredl M, Hofmann F. Continuity between waking activities and dream activities. *Conscious Cogn* 2003; 12 : 298-308.
36. Perogamvros L, Dang-Vu TT, Desseilles M, et al. Sleep and dreaming are for important matters. *Front Psychol* 2013; 4 : 474.
37. Grenier J, Cappeliez P, St-Onge M, et al. Temporal references in dreams and autobiographical memory. *Mem Cognit* 2005; 33 : 280-8.
38. Nielsen TA, Levin R. Dreams, dreaming theories and correlates of nightmares. In : Squire LR, éd. *Encyclopedia of Neuroscience*. Oxford : Academic Press, 2009.
39. Nielsen TA, Kuiken D, Alain G, et al. Immediate and delayed incorporations of events into dreams: further replication and implications for dream function. *J Sleep Res* 2004; 13 : 327-36.
40. Wamsley EJ. Dreaming and offline memory consolidation. *Curr Neurol Neurosci Rep* 2014; 14 : 433.
41. Diekelmann S, Born J. The memory function of sleep. *Nat Rev Neurosci* 2010; 11 : 114-26.
42. Rauchs G, Desgranges B, Foret J, et al. The relationships between memory systems and sleep stages. *J Sleep Res* 2005; 14 : 123-40.
43. Peigneux P, Laureys S, Fuchs S, et al. Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* 2004; 44 : 535-45.
44. Wamsley EJ, Stickgold RJ. Memory, sleep and dreaming: experiencing consolidation. *Sleep Med Clin* 2011; 6 : 97-108.
45. Schredl M, Hoffmann L, Sommer JU, et al. Olfactory stimulation during sleep can reactivate odor-associated images. *Chemosens Percept* 2014; 7 : 140-6.
46. Stickgold RJ, Walker MP. Sleep-dependent memory triage: evolving generalization through selective processing. *Nat Neurosci* 2013; 16 : 139-45.
47. Crick F, Mitchison G. The function of dream sleep. *Nature* 1983; 304 : 111-4.
48. Fischer S, Diekelmann S, Born J. Sleep's role in the processing of unwanted memories. *J Sleep Res* 2011; 20 : 267-74.
49. Baylor GW, Cavallero C. Memory sources associated with REM and NREM dream reports throughout the night: a new look at the data. *Sleep* 2001; 24 : 165-70.
50. Jouvet M. What does a cat dream about? *Trends Neurosci* 1979; 2 : 280-2.
51. Revonsuo A. The reinterpretation of dreams: an evolutionary hypothesis of the function of dreaming. *Behav Brain Sci* 2000; 23 : 877-901.
52. De Koninck J, Christ G, Hébert G, et al. Language learning efficiency, dreams and REM sleep. *Psychiatr J Univ Ott* 1990; 15 : 91-2.
53. Solomonova E, Stenstrom P, Paquette T, et al. Different temporal patterns of memory incorporations into dreams for laboratory and virtual reality experiences: relation to dreamed locus of control. *Int J Dream Res* 2015; 8 : 10-26.
54. Fox KCR, Nijeboer S, Solomonova E, et al. Dreaming as mind wandering: evidence from functional neuroimaging and first-person content reports. *Front Hum Neurosci* 2013; 7 : 412.
55. Walker MP, Liston C, Hobson JA, et al. Cognitive flexibility across the sleep-wake cycle: REM-sleep enhancement of anagram problem solving. *Cogn Brain Res* 2002; 14 : 317-24.
56. Ghosh VE, Gilboa A. What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia* 2014; 53 : 104-14.
57. Maquet P, Péters J, Aerts J, et al. Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. *Nature* 1996; 383 : 163-6.
58. Lewis PA, Durrant SJ. Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn Sci* 2011; 15 : 343-51.
59. Perogamvros L, Schwartz S. The roles of the reward system in sleep and dreaming. *Neurosci Biobehav Rev* 2012; 36 : 1934-51.
60. Desseilles M, Dang-Vu TT, Sterpenich V, et al. Cognitive and emotional processes during dreaming: a neuroimaging view. *Conscious Cogn* 2011; 20 : 998-1008.
61. Tononi G, Cirelli C. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron* 2014; 81 : 12-34.
62. Tamminen J, Lambon Ralph MA, Lewis PA. The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J Neurosci* 2013; 33 : 15376-81.

63. Maquet P, Ruby P. Psychology: insight and the sleep committee. *Nature* 2004; 427 : 304-5.
64. Guérolé F, Marcaggi G, Baleyte J-M, et al. Dreams in normal and pathological aging. *Psychol Neuropsychiatr Vieil* 2010; 8 : 87-96.
65. Chellappa SL, Münch M, Blatter K, et al. Does the circadian modulation of dream recall modify with age? *Sleep* 2009; 32 : 1201-9.
66. Chellappa SL, Münch M, Knoblauch V, et al. Age effects on spectral electroencephalogram activity prior to dream recall. *J Sleep Res* 2012; 21 : 247-56.
67. Nielsen TA. Variations in dream recall frequency and dream theme diversity by age and sex. *Front Neurol* 2012; 3 : 106.
68. Borek LL, Kohn R, Friedman JH. Phenomenology of dreams in Parkinson's disease. *Mov Disord* 2007; 22 : 198-202.
69. Valli K, Frauscher B, Gschiesser V, et al. Can observers link dream content to behaviours in rapid eye movement sleep behaviour disorder? A cross-sectional experimental pilot study. *J Sleep Res* 2012; 21 : 21-9.
70. Olson EJ, Boeve BF, Silber MH. Rapid eye movement sleep behaviour disorder: demographic, clinical and laboratory findings in 93 cases. *Brain* 2000; 123 : 331-9.

Neurosciences et mémoires collectives : les schémas entre cerveau, sociétés et cultures

Nicolas Legrand^{1,2,3,4}, Pierre Gagnepain^{1,2,3,4}, Denis Peschanski⁵ et Francis Eustache^{1,2,3,4,6}

¹ Inserm, U1077, 14000 Caen, France

² Université de Caen Normandie, UMR-S1077, 14000 Caen, France

³ École pratique des hautes études, UMR-S1077, 14000 Caen, France

⁴ CHU de Caen, UMR-S1077, 14000 Caen, France

⁵ CNRS – Université Paris I Panthéon Sorbonne, Centre d'Histoire Sociale du XXème siècle, 75004 Paris, France

⁶ Inserm-EPHE-UNICAEN U1077, Pôle des Formations et de Recherche en Santé (PFRS), Neuropsychologie et neuroanatomie fonctionnelle de la mémoire humaine, 2 rue des Rochambelles, 14032 Caen Cedex, France

Auteur correspondant : Francis Eustache, francis.eustache@unicaen.fr

Reçu le 26 octobre 2015

Résumé – Ces vingt dernières années, l'influence des rapports intersubjectifs sur la cognition a été un sujet d'étude émergent dans les neurosciences cognitives conduisant, à travers ce qui a été décrit comme un « tournant social », à l'apparition de ramifications intégrant sociétés et cultures à ce domaine de recherche. Cette tendance concerne aujourd'hui l'étude de la mémoire collective, définie comme un ensemble de représentations du passé constitutives de l'identité d'un groupe, sélectionnées et partagées par l'ensemble de ses membres connectés par une histoire commune. Après une description de ces évolutions dans l'exploration du cerveau et des comportements, nous proposons une revue des recherches récentes qui ont amorcé la réunion de la psychologie cognitive, des neurosciences et des sciences sociales dans l'étude de la mémoire collective. En utilisant le concept de schéma mnésique, qui a été largement mobilisé dans la neuropsychologie et la neurobiologie de la mémoire, nous proposons un cadre théorique qui permettrait de rendre compte de la formation de souvenirs collectifs en nous concentrant sur les processus d'encodage d'événements historiques. Nous soulignons le fait que l'intégration des recherches sur les bases neurales des schémas mnésiques avec les études sur la mémoire collective pourrait conduire à une meilleure compréhension des interactions réciproques entre la mémoire individuelle et les ressources culturelles tels que les médias ou l'éducation.

Mots clés : Mémoire collective / cognition sociale / ressources culturelles / histoire / schémas mnésiques

Abstract – Neuroscience and collective memory: memory schemas linking brain, societies and cultures.

During the last two decades, the effect of intersubjective relationships on cognition has been an emerging topic in cognitive neuroscience leading through a so-called “social turn” to the formation of new domains integrating society and cultures to this research area. Such inquiry has been recently extended to collective memory studies. Collective memory refers to shared representations that are constitutive of the identity of a group and distributed among all its members connected by a common history. After briefly describing those evolutions in the study of human brain and behaviors, we review recent researches that have brought together cognitive psychology, neuroscience and social sciences into collective memory studies. Using the reemerging concept of memory schema, we propose a theoretical framework allowing to account for collective memories formation with a specific focus on the encoding process of historical events. We suggest that (1) if the concept of schema has been mainly used to describe rather passive framework of knowledge, such structure may also be implied in more active

fashions in the understanding of significant collective events. And, (2) if some schema researches have restricted themselves to the individual level of inquiry, we describe a strong coherence between memory and cultural frameworks. Integrating the neural basis and properties of memory schema to collective memory studies may pave the way toward a better understanding of the reciprocal interaction between individual memories and cultural resources such as media or education.

Key words: Collective memory / social cognition / cultural resources / history / memory schemas

1 Introduction

La description de la mémoire par la neuropsychologie et les neurosciences a connu ces dernières années un « tournant social » qui s'est caractérisé par la prise en compte croissante de l'influence des rapports interpersonnels et de l'environnement culturel sur la formation, le partage et la transformation des souvenirs. Ce tournant atteint aujourd'hui un point déterminant avec le développement d'études portant sur le versant cognitif de la *mémoire collective* (Barnier & Sutton, 2008; Coman *et al.*, 2009a; Anastasio *et al.*, 2012; Roediger & Abel, 2015). Cette tendance signe une rupture épistémique importante car si cette thématique s'était imposée depuis les années 1970, ce fut essentiellement dans le cadre d'une clôture disciplinaire qui opposait largement son ancrage cognitif d'un côté, et culturel et social de l'autre (Olick, 1999; Hirst & Manier, 2008). En particulier, des propositions contraires pouvaient apparaître quant à la nature précise des souvenirs collectifs et aux moyens nécessaires pour leur étude. Une première approche favoriserait ainsi la description de la mémoire à un niveau individuel et psychologique pour la situer ensuite dans son contexte collectif, c'est-à-dire au sein d'un groupe plus ou moins étendu. Une seconde approche au contraire la situerait avant tout dans des ressources sociale et culturelle qui ne se réduisent pas à l'ensemble des interactions entre les membres du groupe, et qui doivent donc être étudiées par des disciplines consacrées, notamment les sciences humaines et sociales (sociologie, histoire, anthropologie...).

Ces deux approches soulignent des conceptions difficilement conciliables des processus qui permettent le passage de niveaux d'explications centrés autour de l'individu à d'autres qui englobent le corps social et culturel dans son ensemble. Pourtant, les nombreuses études récentes sur la mémoire collective ont souligné l'importance et la nécessité d'interactions transdisciplinaires entre la psychologie, les neurosciences et les sciences humaines et sociales autour d'un objet et de méthodes communs. Nous proposons une revue de ces études en nous concentrant sur le concept de *schéma* qui a été largement mobilisé dans ces champs disciplinaires, avec des définitions également variées, et qui réapparaît aujourd'hui fortement dans les études de

neurosciences. En nous fondant sur l'une des premières formulations de ce concept par le psychologue Frédéric Bartlett (1932), nous montrons que celle-ci pourrait permettre de résoudre certaines oppositions entre ces deux approches de la mémoire collective. À travers une définition unique, les schémas permettent en effet de capturer des fonctions, des organisations et des orientations fondamentales de la mémoire dans une dimension tant individuelle que collective. Ces modèles, et leurs applications dans les neurosciences actuelles donnent de nouvelles clés pour penser l'inscription des souvenirs historiques et collectifs dans le cerveau et permettent de reconsidérer les rapports entre des niveaux d'explication psychologiques et sociaux. Mais ils imposent également une évolution du concept de mémoire qui est utilisé par les neurosciences et qui, loin d'être socialement isolable, se trouve structuré par des significations culturelles et collectives profondes.

Le tournant social de la neuropsychologie de la mémoire

La mémoire désigne, dans un sens relativement large, la fonction grâce à laquelle l'homme entretient son rapport au temps. Elle peut renvoyer à des processus psychologiques individuels permettant la conservation dans le cerveau de traces d'expériences passées, mais également à une attitude collective des hommes face à leur histoire en se plaçant sur une échelle qui dépasse celle de la vie des individus. La première définition a longtemps dominé l'étude de la mémoire, à commencer par la philosophie où l'exercice de cette fonction était réduite à l'utilisation de souvenirs par un sujet, tendance que le philosophe Paul Ricoeur a qualifiée de « *tradition du regard intérieur* » (Ricoeur, 2000, p. 115). Un souvenir est considéré comme ce qui peut être saisi par une conscience qui cherche à « *atteindre rétrospectivement toute action ou pensée passée* » (Locke, 1689). Cette tradition a ouvert la voie notamment à une meilleure compréhension de l'interdépendance entre les souvenirs d'un individu, auxquels il est seul à avoir accès, et son identité. En étant garante de la continuité et de la conservation des représentations dans son évolution, la mémoire insère

ainsi le sujet dans une certaine temporalité et participe, à partir d'expériences multiples, à la création par généralisation des concepts et catégories. En revanche, les différentes formes de pratiques sociales de la mémoire ont été peu développées, de même que l'influence des contextes culturels et sociaux sur l'identité et les souvenirs d'un individu. Cette approche fut pourtant longtemps dominante et se poursuit dans le développement des méthodes de la psychologie scientifique naissante. À la fin du XIX^{ème} siècle le psychologue allemand Hermann Ebbinghaus, considéré comme l'un des pionniers de l'étude scientifique de la mémoire, faisait de l'isolement la condition même de la psychologie expérimentale, puisqu'il étudiait la mémorisation de syllabes sans signification. Mais ces contraintes furent également reprises lors du développement des sciences cognitives au milieu du XX^{ème} siècle et jusque dans l'expansion des neurosciences au début des années 1990.

Cette conception a été progressivement remise en cause à mesure qu'il est devenu évident qu'il était impossible de dissocier méthodologiquement un processus psychologique du contexte dans lequel il intervient. La neuropsychologie et les neurosciences connaissent ainsi depuis une quinzaine d'années ce qui pourrait être qualifié de *tournant social* dans l'étude de la mémoire et qui vise précisément à aborder celle-ci non plus à travers les circonstances artificielles d'un isolement du sujet mais dans un contexte, socialement et culturellement déterminé. Cette inflexion a commencé par la prise en compte dans le courant des *neurosciences sociales* de l'influence de fonctions comme la théorie de l'esprit, l'empathie ou les émotions complexes sur les processus cognitifs (Cacioppo & Decety, 2011). Ces évolutions ont participé à définir le cerveau comme un système biologique lui-même inséré dans un univers social dont on pourrait difficilement l'extraire sans que cela interfère sur les processus cognitifs, même à un niveau élémentaire. La volonté de prendre en compte l'influence des inputs sociaux sur la cognition a été étendue ces dernières années bien au-delà de l'étude d'interactions simples entre quelques individus pour englober finalement les notions de groupes et de catégories sociales (Van Bavel & Cunningham, 2012; Cikara & Van Bavel, 2014) ainsi que l'environnement culturel dans son ensemble avec l'émergence des *cultural neurosciences* (Kim & Sasaki, 2014; Han & Ma, 2015). De telles extensions de ce domaine de recherche impliquent des échanges, voire des intégrations transdisciplinaires fortes (Cacioppo & Cacioppo, 2013), notamment avec les sciences humaines et sociales traditionnellement dédiées à ces études (Peschanski, 2013; Peschanski & Maréchal, 2013). Des dialogues entre des univers scientifiques historiquement distants ont ainsi émergé ces dernières années, comme en témoignent l'apparition de « trans-disciplines »

comme la *neurosociologie* (Franks & Turner, 2013), la *neuroanthropologie* (Dominguez Duque *et al.*, 2010) ou encore la *neurohistoire* (Smail, 2008; Mandressi, 2011; Burman, 2012).

Cette évolution du domaine de la neuropsychologie et des neurosciences se répercute également sur la définition même qu'il convient de donner à la mémoire, qui a intégré progressivement l'influence du cadre social. Cette idée découle assez largement des travaux du sociologue Emile Durkheim mais surtout par la suite de son élève Maurice Halbwachs qui publia en 1925 *Les cadres sociaux de la mémoire*, considéré aujourd'hui comme l'ouvrage de référence dans l'approche sociale de cette fonction. Halbwachs s'opposait alors pour partie au philosophe Henri Bergson qui avait abordé la mémoire dans une perspective essentiellement phénoménologique, et soulignait à l'inverse que « *c'est dans la société que, normalement, l'homme acquiert ses souvenirs, qu'il se les rappelle, et, comme on dit, qu'il les reconnaît et les localise* » (Halbwachs, 1926, p. VI). Aujourd'hui, l'idée selon laquelle le fonctionnement d'une mémoire saine s'imprègne nécessairement du langage, d'interactions interpersonnelles de catégories et de pratiques culturelles, est encore d'actualité et a été abondamment décrite par la suite dans les sciences humaines et sociales.

À la suite de cette notion de cadres sociaux de la mémoire, la mémoire collective a été définie comme un ensemble de représentations, partagées sur le long terme par les membres d'un groupe connectés par un passé commun, et qui participent à la constitution de son identité (Hirst & Manier, 2008). Cela implique pour les neurosciences de comprendre comment peuvent émerger au sein d'un groupe des représentations partagées, et comment l'environnement social et culturel peut influencer et façonner le contenu même de la mémoire individuelle. Cette mémoire est paradoxalement issue d'expériences et d'apprentissages uniques qui sont propres à chaque personne, mais également d'une réalité commune (Echterhoff *et al.*, 2009), ce qui impliquerait que des structures ou des éléments de ces souvenirs devraient être communs. L'encodage d'événements historiques est un exemple central à la fois de ce partage de représentations identiques au sein de « communautés mnésiques » (*mnemonic community*, Hirst & Manier, 2008) et de l'influence des représentations culturelles dans l'établissement de ces souvenirs.

L'intérêt de la psychologie cognitive pour la mémoire collective a grandi ces dernières années (Manier & Hirst, 2007; Barnier & Sutton, 2008; Hirst & Manier, 2008; Sutton *et al.*, 2008; Anastasio *et al.*, 2012; Brown *et al.*, 2012a; Coman *et al.*, 2009a; Hirst & Rajaram, 2014; Roediger & Abel, 2015). Dans une dimension moindre, l'appropriation de ce concept par

les sciences cognitives s'est également répercutée dans les neurosciences (Edelson *et al.*, 2011; Kim & Sasaki, 2014; Han & Ma, 2015). Pourtant, l'intrication des processus cognitifs avec des niveaux sociaux et culturels reste difficile à établir clairement. Cette difficulté s'explique assez largement par la polysémie même du concept de mémoire collective. Les travaux de Halbwachs prenaient déjà en compte l'état des connaissances de la psychologie, de la philosophie, de la psychanalyse et de la sociologie naissante, mais ont été poursuivis dans des univers disciplinaires éloignés qui n'ont pas toujours reproduit le mouvement d'intégration initial. Il existe ainsi une opposition récurrente, bien qu'elle ne soit pas toujours clairement formulée, entre l'étude d'une mémoire collective à partir de l'évolution et de la communication de souvenirs par des individus, et une mémoire apparaissant et se déterminant au niveau social (Zerubavel, 1996) ou culturel (Assmann, 2008) dont le fonctionnement ne se réduirait pas aux processus observables à un niveau individuel (Olick, 1999). Ces divergences peuvent par exemple conduire à localiser différemment les souvenirs collectifs : dans le cerveau pour la psychologie ou dans le monde extérieur et les ressources culturelles pour les sciences humaines et sociales (Coman *et al.*, 2009). De fait, en fonction des disciplines, le concept de mémoire collective peut s'attacher à différents processus impliqués dans la formation de souvenirs collectifs qu'il convient de distinguer.

Mémoires collectives : partage, diffusion et influence des représentations sociales

Le débarquement allié du 6 juin 1944 sur les plages normandes est un exemple représentatif de ces souvenirs historiques qui se trouvent profondément ancrés dans les mémoires collectives et ce, au-delà même des frontières françaises. La constitution de ces événements-monde et de ces lieux-monde ne fut pas nécessairement immédiate et, comme nous le verrons, l'ensemble des données historiques ne remplirent pas les conditions de la mise en récit mémorielle. Pourtant les processus d'écriture ont commencé très tôt. La première approche de la mémoire collective que nous pouvons alors évoquer ici décrit (*cf.* Mécanisme 1; Figure 1) la pratique sociale de la mémoire des témoins, qui ont pu partager ou occulter à propos de l'événement certaines dimensions de leurs souvenirs à travers les communications et discussions menées entre eux mais également avec les générations suivantes. Cette mise en commun d'informations avec les autres membres du groupe est une étape déterminante dans la dimension collective et historique de la mémoire. Elle se distingue toutefois, dans un premier temps, de l'idée de *souvenirs collectifs*

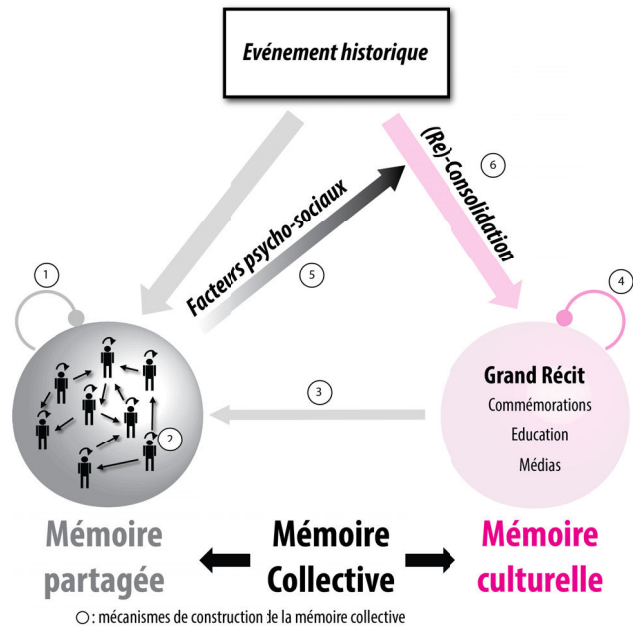


Fig. 1. Mécanismes de construction de la mémoire collective. L'inscription d'un événement dans la mémoire collective engage plusieurs mécanismes psychologiques et sociaux. 1. Le partage des souvenirs entre les individus, permettant leur diffusion dans le groupe et l'émergence de représentations prototypiques et de schémas. 2. Des processus individuels de sélection de certains souvenirs influencés par l'état des connaissances préexistantes en mémoire. 3. Ces connaissances préexistantes peuvent être issues des ressources culturelles qui exercent une influence sur la mémoire individuelle. 4. Des mécanismes de récupération et de consolidation de l'événement peuvent également apparaître à un niveau social à travers des politiques de valorisation des mémoires. 5. Des facteurs psycho-sociaux comme le besoin de communiquer l'événement ou la possibilité pour les générations en ayant souffert d'instaurer des lieux de conservation de la mémoire peuvent également intervenir. 6. L'apparition d'un événement important peut également induire la re-consolidation d'autres événements plus anciens qui lui sont liés.

qui implique que des représentations identiques se trouvent encodées par l'ensemble des membres du groupe, et de *mémoire culturelle* qui renvoie à un ensemble externalisé de connaissances. C'est toutefois à partir du témoignage « *Structure fondamentale de transition entre la mémoire et l'histoire* » (Ricoeur, 2000, p. 26), qu'émergent par la suite des représentations synthétiques et générales à propos d'événements vécus en commun.

La mise en récit d'événements est une fonction essentielle du langage (Dessalles, 2007) et l'évocation collective qu'elle induit, ou souvenir interactif (*reminding*), se distingue de formes plus classiques comme

le simple rappel (*reminding*) d'une part ou la reconnaissance (*recognizing*) d'autre part (Casey, 1987). « *Le plus souvent* », notait déjà Halbwachs (1925), « *si je me souviens c'est que les autres m'incitent à me souvenir, que leur mémoire vient au secours de la mienne, que la mienne s'appuie sur la leur* ». Les mécanismes psychologiques liés au rappel collaboratif et à l'influence des interactions interindividuelles sur la mémoire personnelle ont été l'objet de nombreuses études en psychologie sociale et cognitive (Coman & Hirst, 2012; Hirst & Echterhoff, 2012; Stone & Hirst, 2014; Stone *et al.*, 2010, 2013). Il est apparu en particulier que les individus rappellent plus d'informations en groupe que s'ils sont seuls (Barnier *et al.*, 2008; Harris *et al.*, 2014), mais que le rappel collaboratif du groupe reste toutefois inférieur à la somme des souvenirs à disposition des différents membres. Seules certaines informations sont évoquées : celles qui sont partagées, pertinentes et socialement valorisables. Des souvenirs, initialement variés et individuels, tendent ainsi à une certaine uniformisation lorsqu'ils sont communiqués pour former une mémoire partagée (Roediger *et al.*, 2009). Les discussions peuvent ainsi être un vecteur de modification des souvenirs tant pour l'auditeur que pour le narrateur (Cuc *et al.*, 2006), la mémoire individuelle offrant une grande malléabilité, notamment sous l'influence de facteurs sociaux (Edelson *et al.*, 2011).

Mais au-delà de cette pratique sociale de la mémoire, la mémoire collective peut aussi être approchée à travers l'ensemble des représentations significatives et constitutives de l'identité du groupe qui les partage (*cf.* Mécanisme 2; Figure 1). La mémoire collective doit être envisagée en lien étroit avec la notion d'identité collective (*e.g.* Lavabre, 1994). Un souvenir collectif est davantage qu'un souvenir partagé, qui peut aussi bien impliquer des connaissances générales sans portée sociale. Les représentations partagées en France autour du débarquement dépassent la simple description factuelle de l'événement et l'insèrent dans une description cohérente de l'identité nationale. À l'image du souvenir individuel qui est à l'origine de l'identité du sujet, le souvenir collectif doit être un élément de l'identité collective (Manier & Hirst, 2007) indiquant comment le groupe est devenu ce qu'il est actuellement, quelles en sont ses valeurs mais aussi à quels groupes ils s'opposent. Cette définition se rapproche alors fortement de l'étude du sentiment d'appartenance à un groupe ou à une catégorie sociale qui a récemment émergé dans le domaine des neurosciences (pour revue voir Cikara & Van Bavel, 2014). De fait, le partage de souvenirs collectifs est un facteur facilitant le sentiment d'appartenance au même groupe (Tavani *et al.*, 2015). Mais à la question de l'interdépendance entre identité collective et souvenirs individuels s'ajoutent également les mécanismes

de sélection des représentations les plus significatives et le processus de mise en récit mémorielle. Que ce soit dans l'expérience directe d'un événement ou dans le processus d'écriture de l'histoire, les mêmes phénomènes ne revêtent pas la même importance selon les cultures (Pennebaker *et al.*, 2006). La question des propriétés nécessaires à un événement pour son inscription en mémoire est une question partagée par l'histoire (Dosse, 2010; Peschanski, 2012) et la psychologie (*e.g.* Rubin & Umanath, 2015) qui tentent ainsi de définir les caractéristiques et les propriétés des événements auxquelles la mémoire humaine est la plus sensible.

Pourtant la mémoire collective ne se limite pas à un ensemble de représentations que partagent les membres d'un groupe mais peut aussi désigner des éléments qui sont présents dans l'environnement culturel du sujet et qui témoignent de son histoire (*cf.* Mécanisme 3; Figure 1). Les concepts centraux ici de temps, d'identité et de mémoire peuvent toujours se décrire à la fois sur des niveaux personnels, sociaux et culturels ayant chacun leurs modalités propres (Assmann, 2008). Il importe alors de distinguer une mémoire communicative que nous qualifierons ici de *mémoire partagée*, qui correspond aux souvenirs collectifs pouvant être transmis à travers des échanges directs, et une *mémoire culturelle* qui désigne l'externalisation de ces souvenirs dans des ressources dédiées de façon à en assurer la pérennité. La mémoire collective associée à la seconde guerre mondiale s'est par exemple constituée autour de vecteurs comme la radio, la télévision, le cinéma, la littérature, les enseignements scolaires, les musées ou même aujourd'hui les réseaux sociaux. Ces ressources influencent de façon déterminante le contenu et la structure même des souvenirs individuels. Pourtant, si les niveaux personnels et communicatifs de la mémoire ont fait l'objet de nombreux travaux en psychologie et en neurosciences, la dimension culturelle n'a été abordée que récemment.

La mémoire collective peut également être pointée comme pratique sociale (*cf.* Mécanisme 4; Figure 1). Les notions de « devoir de mémoire » (Ledoux, 2012) ou encore de « concurrence des mémoires » (Chaumont, 2002) montrent aujourd'hui que la conservation du souvenir n'est pas un processus automatique mais s'accompagne de débats sur l'interprétation qu'il convient de donner du passé (voir aussi Ricoeur, 2000). Son inscription dans un récit qui lui donne sens relève également de volontés sociales et politiques qui s'expriment à travers cérémonies, commémorations, reconnaissances publiques ou encore constructions de mémoriaux qui arrachent certaines traces du passé à l'oubli pour en faire ce que l'historien Pierre Nora qualifie de *lieux de mémoire* (Nora, 1984). Cette dernière dimension relève encore

assez largement de politique et de sociologie et sera moins considérée ici. Enfin, la réactivation d'un souvenir collectif peut être influencée par des facteurs psychosociaux (*cf.* Mécanisme 5 ; Figure 1) comme la résonance émotionnelle de celui-ci. Ces facteurs induiront des processus de partage ou de répression de l'événement qui vont moduler sa reconsolidation dans les ressources culturelles (*cf.* Mécanisme 6 ; Figure 1). Toutefois, ces aspects ne seront pas abordés dans le cadre de cette revue.

Les processus de communication des souvenirs au sein d'un groupe (1), leur sélection et leur rôle dans la construction de l'identité collective (2) et l'influence des ressources culturelles sur la mémoire individuelle (3) sont des domaines encore largement explorés par les neurosciences cognitives. Il est pourtant indéniable que des mécanismes cognitifs et neuronaux interviennent de façon déterminante dans ces processus. Nous suggérons ici que la création et le maintien de représentations collectives renvoient à ce que la psychologie cognitive et les neurosciences ont rassemblé sous le concept de *schémas*. Ce concept, proposé au début du XX^{ème} siècle (Head & Holmes, 1911), a traversé de nombreuses disciplines, de la neurologie aux sciences sociales, en permettant de penser à chaque fois de façon différente une dimension essentielle de la mémoire humaine mais également l'inscription de celle-ci au sein de pratiques culturelles. Nous proposons ici un cadre théorique issu d'études neuroscientifiques récentes permettant de décrire l'encodage de souvenirs collectifs historiques ainsi que leur influence dans la construction de la mémoire individuelle.

Parallèle entre schémas mnésiques et schémas culturels

La mémoire humaine est une fonction qui semble prise entre deux impératifs contradictoires : celui d'une rigidité absolue qui permettrait aux représentations anciennes d'être conservées dans le temps, et celui d'une grande flexibilité pour que de nouvelles informations intègrent constamment les représentations préexistantes afin que celles-ci se coordonnent à l'environnement. Il est établi sur ce point que l'hippocampe est dédié à l'encodage d'événements spécifiques, et le néocortex à l'intégration, plus lente, des régularités de l'environnement qui ne disposent pas encore de schémas. L'approche neurobiologique de la mémoire a montré que les traces mnésiques, loin d'être rigides, traversent des phases successives de consolidation, fragilisation, transformation et reconsolidation en fonction de nombreux facteurs comme leur utilisation ou leur importance (Tronson & Taylor, 2007). Mais l'intégration de nouvelles informations est un processus complexe reposant sur la plasticité cérébrale et qui peut potentiellement affecter

les traces mnésiques plus anciennes. Le concept de schéma est devenu un élément central en permettant de répondre à ces deux exigences, essentiellement grâce à la définition proposée par le psychologue britannique Frédéric Bartlett (1932). Pour ce dernier, un schéma désigne ainsi un ensemble organisé de représentations issues d'expériences passées qui jouent un rôle de régulateur et de filtre face aux expériences nouvelles et à l'anticipation des expériences futures. Mais cette notion a largement débordé le cadre de la psychologie cognitive et la même analogie peut être faite en considérant non plus une structure mentale et cérébrale, mais une construction culturelle, sociale et historique cristallisée sur des supports différents (médias, éducation, livres, Mémoires, *etc.*) qui obéissent néanmoins aux mêmes règles de fonctionnement (Wagoner, 2012, 2013a). Pourtant, avant l'étude princeps de Tse *et al.* (2007), cette idée fondamentale pour la mémoire a eu peu d'implications pour les neurosciences.

Structure cognitive

La mémoire favorise donc l'émergence de ces structures qui vont organiser de façon dynamique les informations encodées. Une expérience n'apparaît jamais de façon isolée à la conscience mais au premier plan d'un ensemble structuré d'expériences passées (« *whole active mass of organized past reactions or experience* » Bartlett, 1932b, p. 213) qui conditionnent les réponses de l'individu. Cette définition est un jalon important de la psychologie de la mémoire puisqu'elle a souligné de façon déterminante que l'évocation d'un souvenir n'a pas pour fonction d'accéder à l'expérience telle qu'elle a été vécue initialement mais d'en fournir une reconstruction cohérente avec le contexte actuel, l'action du sujet et l'anticipation du futur à laquelle il se livre.

Les neurosciences définissent aujourd'hui les schémas comme des réseaux associatifs de connaissances, adaptables et héritées de multiples expériences dont les détails ont été effacés afin de n'en conserver que les éléments essentiels (Ghosh & Gilboa, 2014), par exemple pour mémoriser la disposition des objets dans une pièce qui est très familière. Toutefois, les schémas se distinguent de structures plus abstraites comme les concepts ou les catégories car en tant que synthèses d'expériences ils en conservent encore certains aspects et sont surtout susceptibles d'évoluer en fonction de situations nouvelles. Cette fonction d'adaptation (et son rôle dans la mémorisation) est un élément central dans leur description par les neurosciences. L'étude princeps de Tse *et al.* (2007) a en effet montré que les nouvelles informations deviennent plus rapidement indépendantes de l'hippocampe si elles sont intégrées à des schémas

préexistants. Toutefois, si une information doit être cohérente avec les représentations précédentes, elle doit également en différer de manière significative pour que sa mémorisation soit utile à l'individu, ce qui implique que la nouveauté (en relation avec le schéma préexistant) soit également une composante pouvant favoriser l'intégration de nouvelles informations (Henson & Gagnepain, 2010; van Kesteren *et al.*, 2012). Depuis, de nombreuses études ont souligné l'importance d'un réseau reliant certaines parties du lobe temporal médian, dont l'hippocampe, avec le cortex préfrontal médian dans l'utilisation et la modification des schémas en fonction des besoins liés à certaines épreuves de mémoire, chez le rat (Tse *et al.*, 2011) et chez l'Homme (van Kesteren *et al.*, 2010). Ces réseaux clés permettent l'intégration d'informations avec les structures qui reflètent les propriétés récurrentes de l'environnement (van Kesteren *et al.*, 2013). Mais les études en neuropsychologie ont également montré que le trouble le plus fréquemment associé à une lésion du cortex préfrontal médian est la confabulation. Une étude récente (Ghosh *et al.*, 2014) a établi que de telles lésions n'altèrent pas simplement les processus d'intégration de nouveaux souvenirs propres aux schémas, mais affectent de manière générale tout processus cognitif les impliquant, comme par exemple le fait de devoir catégoriser des mots en fonction de leur lien ou non avec des actions quotidiennes (« *aller au lit* » ou « *aller chez le docteur* »), tâches engageant les schémas associés.

Ces résultats suggèrent que ces structures ne se limitent pas à une fonction de facilitation de l'intégration de l'information, mais interviennent constamment dans l'interaction avec l'environnement et peuvent être utilisées de façon active dans des processus de compréhension et d'interprétation. De fait, la mémoire ne se résume pas à la conservation d'éléments récurrents, mais implique que l'individu puisse donner un sens aux expériences auxquelles il est confronté dans le présent (« *effort after meaning* », Bartlett, 1932, p. 227). Cette nécessité s'exprime à un niveau mnésique par le besoin de relier l'information entrante à d'autres éléments déjà présents en mémoire pour ramener la plus grande partie possible de ce qui est nouveau et inconnu à une expérience passée pour en comprendre l'évolution. Un tel processus d'interprétation implique un travail de comparaison et de mise en perspective de schémas distants et peut s'apparenter à la théorie du codage prédictif récemment introduite dans les neurosciences (Friston, 2010; voir également Henson & Gagnepain, 2010). Cette théorie vise à unifier le fonctionnement cérébral sous un seul et même principe et propose que le cerveau génère constamment un modèle interne assimilable à un schéma visant à anticiper les entrées sensorielles extérieures. Lorsque ces prédictions sont

invalidées, un signal d'erreur est alors généré et peut potentiellement déclencher l'incorporation de cette information au modèle interne.

Ce que nous avons décrit précédemment comme l'échange répété autour de souvenirs communs (Mécanisme 1; Figure 1) et l'influence des contextes sociaux et culturels (Mécanisme 3; Figure 1) sont autant de facteurs susceptibles de favoriser l'émergence de structures semblables en mémoire qui rendent compte de la possibilité de partage de représentations et des influences possibles de l'environnement social. Pourtant l'intégration de souvenirs collectifs et historiques à la mémoire collective implique une référence à l'identité du groupe et une signification spécifique (Mécanisme 2; Figure 1) que nous allons maintenant détailler.

Souvenirs historiques et schémas culturels

Les souvenirs historiques renvoient à des événements ayant eu un impact au niveau collectif, et constituant une étape importante dans le développement de l'identité du groupe. De ce fait, l'encodage de ces informations en mémoire implique à la fois des éléments épisodiques relatifs à l'événement lui-même, et une dimension sémantique permettant de décrire et de comprendre des phénomènes qui dépassent le cadre de l'expérience personnelle en les liant à d'autres contenus mnésiques. Un souvenir pourrait difficilement être dissocié de structures impliquées dans sa compréhension et des liens qu'il entretient avec d'autres souvenirs. Le contexte du souvenir, les schémas auxquels il s'associe, sont alors déterminants dans son identification (*i.e.* à quel événement se rapporte le souvenir?), la mise en évidence de structures internes (*i.e.* quels liens ces faits entretiennent-ils entre eux et avec d'autres faits déjà connus?) et leur congruence avec une identité collective (*i.e.* est-ce que ce souvenir révisé ou renforce la représentation du groupe auquel on appartient?). Cette dimension s'avère essentielle dans le cas de la mémoire collective qui ne consiste pas simplement en un partage de représentations identiques mais aussi – de façon déterminante – dans le partage d'un sens donné à ces représentations. Les souvenirs collectifs liés au débarquement ne se limitent pas à un ensemble d'images partagées par les français. Ils impliquent que la signification et la valeur qui est attribuée à cet événement soient également partagées, et peuvent différer des significations et valeurs partagées par exemple dans d'autres pays. Cet événement n'a pas été lié aux mêmes schémas en fonction du groupe d'appartenance des individus, de leur rôle ou de leur histoire personnelle. Un même fait historique peut être interprété de façon très différente selon les opinions, l'appartenance au groupe social ou encore la

possibilité de mettre l'événement en lien avec d'autres événements qui lui sont semblables. Cette fonction d'intégration du souvenir au sein de structures qui en révèlent le sens est particulièrement soulignée par l'approche de la mémoire collective en sciences humaines (*e.g.* Ricoeur, 2000; Barash, 2006).

Le besoin d'attribution de sens conduit à établir des rapports de détermination et de causalité entre les événements, mais aussi de responsabilités, d'intentions et de motifs lorsqu'ils impliquent des actions humaines. Les souvenirs historiques s'insèrent ainsi dans une chronologie d'actions dont la signification est révélée par le récit qui en est fourni et qui ordonne la mémoire collective (Wertsch, 2008). La place et l'importance d'un souvenir sont donc tributaires de la cohérence qu'il peut établir avec les souvenirs d'autres événements qui l'ont précédé, mais également du poids qu'il a dans la détermination des situations présentes. C'est ainsi qu'il a pu être difficile de donner un sens collectif aux bombardements alliés sur les villes de Normandie. Une telle représentation, sans nécessairement être complètement effacée – au moins dans l'immédiat – se trouvait pourtant mise à l'écart du schéma collectif qui fait du débarquement le symbole d'une libération (Knapp, 2014). De la même façon, un événement distant géographiquement ou temporellement perd progressivement de son impact pour l'environnement direct du sujet et ainsi de la proportion qu'il occupe dans la mémoire collective (Pennebaker *et al.*, 2006; Pennebaker & Gonzales, 2009). L'importance d'un souvenir historique ne s'explique donc pas simplement par la nature de celui-ci, mais est fonction de son utilisation dans la compréhension des situations présentes.

L'appartenance à un groupe ou à une culture peut également être déterminante dans la conservation de ces schémas culturels. Nous sommes constamment influencés par les groupes auxquels nous appartenons, qui peuvent sensiblement altérer nos décisions (Cikara & Van Bavel, 2014) ou nos souvenirs (Edelson *et al.*, 2011). De la même façon, la conservation du souvenir d'événements historiques mondiaux et leur hiérarchie sont largement tributaires de la nationalité et de la culture des individus (Pennebaker *et al.*, 2006; Liu *et al.*, 2012). Cette influence ne s'explique pas uniquement par des différences de récits historiques en fonction des cultures. Le souvenir d'un événement n'est pas conservé de la même façon en fonction de la responsabilité du groupe ou de la catégorie sociale à laquelle appartient l'individu. Cela a été observé par exemple dans le cas d'exactions et d'actes moralement condamnables (Coman *et al.*, 2014). Dans ces situations, les individus ont tendance à davantage vouloir légitimer une action moralement condamnable si elle a été commise par un membre de leur propre groupe relativement à un membre d'un groupe opposé. Or ce

processus de légitimation et d'explication implique de lier un souvenir particulier à des représentations permettant de comprendre le comportement du groupe, ce qui a également pour conséquence de renforcer en mémoire ce souvenir et les souvenirs qui permettent de le justifier. Ces mécanismes de justification pourraient également s'étendre aux actions auxquelles l'individu participe directement, le cas des *Einsatzgruppen* étant souvent examiné en histoire, où des souvenirs permettant le maintien d'une certaine légitimité au regard de la morale collective prévalant alors sont également indispensables et renforcés.

Les schémas qui sous-tendent ces mécanismes d'attribution de sens sont, dans une large mesure, hérités de l'expérience directe du sujet et reflètent l'impact de son environnement autant que les processus biologiques innés structurant la mémoire humaine. Mais la compréhension d'événements relevant d'une dimension collective et historique ne se fait pas qu'à travers des schémas acquis à un niveau individuel, mais implique également des représentations culturelles, dont l'influence a été précédemment décrite (Mécanisme 3; Figure 1). Ces ressources renvoient souvent à des représentations d'origines plus anciennes que les souvenirs acquis à l'échelle d'une vie humaine, et qui peuvent être partagées dans un groupe (Assmann, 2008). L'éducation, les médias ou les musées organisent également leurs contenus autour de structures signifiantes, les schémas mnésiques sont ainsi à l'individu ce que les schémas culturels sont au groupe social (Bartlett, 1932; Wagoner, 2013b). Souvenirs individuels et représentations culturelles, bien que semblant désigner des contenus hétérogènes, renvoient pourtant à deux états de la mémoire collective entre lesquels les transferts sont constants et qui participent à leur constitution réciproque. C'est en effet parce que schémas mnésiques et schémas culturels peuvent échanger et communiquer que nous pouvons observer la mise en récit de souvenirs individuels dans la mémoire collective, tout autant que l'effet structurant de la mémoire collective sur les souvenirs individuels à travers les ressources culturelles.

Celles-ci apparaissent en effet comme une source majeure de transmission des mémoires auprès des jeunes générations, davantage même que ne le sont la famille ou les proches ayant directement vécu l'événement (Stone *et al.*, 2014). Les schémas qui sont alors mobilisés dans la compréhension d'événements présents n'ont souvent pas été dérivés d'expériences directes mais ont au contraire été acquis par le biais d'échanges avec des contenus culturels. Les mécanismes cognitifs qui pourraient sous-tendre l'émergence, la stabilisation et éventuellement l'oubli des représentations historiques sont aujourd'hui au centre de recherches croisant la psychologie sociale et les sciences cognitives (Pennebaker *et al.*, 2006;

Glăveanu & Yamamoto, 2012; Jovchelovitch, 2012; Knights, 2012). Celles-ci avancent souvent l'idée selon laquelle l'organisation des souvenirs et du récit historique pourrait être décelée dans l'organisation de la mémoire individuelle (Brown *et al.*, 2009). Le concept de schéma montre que, s'il est possible d'observer des mécanismes biologiques et cognitifs universels dans l'organisation des souvenirs, le cerveau est également sensible aux schémas présents dans les ressources culturelles comme autant de structures acquises hors de l'expérience quotidienne et pourtant activement utilisées dans la compréhension de situations nouvelles. On peut par exemple essayer de comprendre les Printemps arabes en les mettant en perspective, au moins temporairement, avec la Révolution française dont la représentation est elle-même organisée autour de schémas récurrents. À l'irréductibilité de l'événement se substitue alors la réémergence par similitudes d'autres représentations directement tirées du récit collectif.

Les schémas culturels et l'oubli collectif

Nous avons suggéré que les schémas mnésiques peuvent intervenir dans l'émergence de souvenirs partagés (Mécanisme 1; Figure 1), l'interprétation d'événements récents et leur inscription dans un récit collectif (Mécanisme 2; Figure 1), et l'interaction entre la mémoire individuelle et les ressources culturelles (Mécanisme 3; Figure 1). Le cadre de recherche transdisciplinaire qui s'est développé ces dernières années s'est toutefois concentré sur les mécanismes qui permettent la conservation et la consolidation des souvenirs collectifs (Anastasio *et al.*, 2012). Pourtant toute représentation du passé n'est pas toujours souhaitée par les sociétés et les individus qui en héritent. Il apparaît au contraire que la mémoire, individuelle comme collective, est contrainte de façon importante par une fonction d'oubli (Eustache *et al.*, 2014). La fonction et les modalités de l'oubli ont été relativement ignorées par les sciences sociales dans l'étude de la mémoire collective (Lavabre, 2013; voir aussi Nietzsche, 1874; Ricoeur, 2000). Il s'agit pourtant d'une dimension essentielle. Si la mémoire collective renvoie à une volonté partagée de se souvenir d'éléments constitutifs d'une forme d'identité, les groupes humains peuvent également partager la volonté de ne pas se souvenir ou du moins de ne pas faire réapparaître certains souvenirs et d'en inscrire d'autres dans leurs ressources culturelles. En neuropsychologie, la mémoire et l'oubli sont parfois présentés comme les deux versants d'une distinction conceptuelle, or on oppose ici en réalité une fonction psychologique (*i.e.* la mémoire) à l'une de ses instances de sélection interne. L'oubli est en effet indissociable

du fonctionnement normal de la mémoire et peut intervenir à différents niveaux. Il peut s'agir (1) d'une suppression volontaire et consciente de certains souvenirs lorsque ceux-ci sont trop difficiles et nuisent au bien-être de l'individu, (2) du résultat d'un accès aux traces mnésiques qui, si elles sont réactivées mais non rappelées peut conduire à leur affaiblissement ou (3) d'une réorganisation et d'une réappropriation du passé dans le présent.

Les processus d'inhibition mnésique sont aujourd'hui largement étudiés par les neurosciences (*e.g.* Anderson & Green, 2001; Gagnepain *et al.*, 2014). Ces études ont mis en avant un système d'inhibition au niveau du cortex préfrontal en lien avec l'hippocampe qui permet de supprimer, de façon volontaire, les traces indésirables. Or, cette suppression d'expériences négatives apparaît à un niveau individuel mais également collectif. Certaines représentations du passé ne sont pas toujours valorisables dans le présent, comme l'a par exemple montré Henry Rousso avec l'évolution des mémoires du régime de Vichy en France (Rousso, 1990). La négativité des représentations ne se fait alors pas uniquement en fonction de l'identité du sujet mais aussi du groupe et de ses valeurs. Si les mécanismes qui induisent à un niveau individuel un besoins d'inhibition de certains souvenirs commencent à être bien décrits par les neurosciences, entrant également en résonance avec une partie du discours psychanalytique (Berlin, 2011), l'influence des catégories sociales et culturelles reste encore mal comprise. Les neurosciences se sont pourtant récemment intéressées aux facteurs d'influence de la cognition comme l'appartenance ou l'opposition à certains groupes sociaux (Cikara & Van Bavel, 2014). De futures études pourront sur ce point analyser au niveau de l'individu l'effet de l'héritage social et culturel sur les besoins d'inhibition de certains souvenirs.

La seconde forme d'oubli qui a été particulièrement étudiée par les neurosciences cognitives est l'oubli induit par la récupération (*retrieval induced forgetting*). Anderson *et al.* (1994) ont montré dans une étude princeps que le simple fait d'accéder à une information peut provoquer une altération des traces d'autres informations qui lui sont liées sémantiquement et conduire à un oubli progressif. Cet effet a été observé isolément, mais a aussi récemment été adapté dans une version sociale dans le cadre d'échanges discursifs (Coman *et al.*, 2009; Brown *et al.*, 2012; Coman & Hirst, 2012; Stone *et al.*, 2012, 2013). Ce partage social de l'oubli induit par la récupération (*socially shared retrieval-induced forgetting*) a alors montré que l'évocation d'une information lors d'une discussion peut aussi induire l'affaiblissement d'éléments qui sont liés à celle-ci et qui ne sont pas évoqués (Cuc *et al.*, 2007; Hirst & Echterhoff, 2012), et ainsi participer à la formation de représentations prototypiques.

Coman *et al.* (2009) ont par exemple montré, en demandant à des participants new-yorkais d'évoquer ensemble leurs souvenirs du 11 septembre 2001, que des discussions pouvaient induire des difficultés d'accès aux éléments propres à l'expérience de chaque individu s'ils n'étaient pas mentionnés, favorisant progressivement l'émergence de représentations uniques au détriment de détails moins signifiants pour le groupe. D'une façon assez paradoxale, l'oubli intervient davantage pour des informations qui ne sont pas mentionnées lors d'une discussion que si la discussion n'avait pas eu lieu (Stone & Hirst, 2014).

La troisième forme d'oubli correspond à une adaptation du passé pour le conformer au présent et peut également intervenir à un niveau plus social. La fonction de la mémoire ne se résume pas à la conservation des traces d'événements passés, mais s'oriente constamment en fonction de leur utilisation future. Par exemple, à un niveau cérébral, les zones impliquées dans la récupération des souvenirs sont dans une large mesure superposables avec les zones responsables de l'anticipation ou de la projection mentale dans le futur (Schacter *et al.*, 2007; Viard *et al.*, 2011). Or l'influence du contexte et des motivations du rappel sur l'organisation et la valorisation des souvenirs pourrait également se retrouver à un niveau social. La conservation d'un souvenir n'est pas simplement fonction de la part du passé de la société qu'il permet de comprendre, mais aussi de son potentiel de résonance avec ses intérêts, présents et futurs. La référence aux heures les plus sombres de l'Histoire, lorsqu'elle intervient dans le domaine social, ne se fait pas dans la simple perspective d'un devoir de souvenir ou d'une volonté de préservation des traces. Elle se voit immédiatement insérée dans une projection vers l'avenir comme autant de leçons sur les écueils de l'humanité, si ce n'est comme contre-argument social ou idéologique. La référence à l'assertion selon laquelle celui qui ne connaît pas l'histoire est condamné à la revivre peut aussi être mise en perspective avec les études comportementales qui montrent à un niveau individuel que c'est lorsque l'on s'expose à une réapparition du passé que son souvenir est nécessaire.

Les scénarios que nous envisageons et les comportements que nous anticipons en fonction de nouvelles expériences peuvent favoriser la consolidation de certaines informations (*e.g.* Wilhelm *et al.*, 2011) mais aussi transformer des souvenirs déjà inscrits dans le cerveau (*e.g.* Dunsmoor *et al.*, 2015). De la même façon, les différentes projections politiques ou sociales que peut se proposer une société en fonction d'événements récents mobilisent chacune des références historiques qui leur sont propres et peuvent ainsi influencer la force de certaines représentations dans la mémoire collective. Les souvenirs les plus valorisés et réinvestis dans le récit historique ou les plus

fréquemment réactivés lorsqu'ils se lient par similitude aux événements récents tendront ainsi à être augmentés relativement aux informations passées sous silence. On pourrait alors observer, à l'image de la mémoire individuelle, une restructuration constante des représentations du passé en mémoire collective pour correspondre aux besoins du présent (Olick & Robbins, 1998) et aux horizons d'attente.

Il est encore difficile de déterminer si cette forme de sélection est un vecteur important de l'oubli à un niveau collectif et s'il est possible de savoir quels souvenirs y sont le plus exposés. La transformation de la mémoire semble opérer à la fois dans la valorisation dans le présent de certaines représentations au cœur d'enjeux sociaux ou historiques, réinvesties, reconstruites et donc transformées pour interpréter les événements actuels. Elle apparaît également lorsqu'au contraire certaines représentations sont écartées ou dévalorisées car elles ne donnent pas à comprendre les composantes essentielles de l'identité collective présente et l'insertion de celle-ci dans une histoire plus large. L'effet d'un tel passage sous silence des mémoires reste encore assez mal compris et de futures études pourront en évaluer l'implication, tant sur les souvenirs présents à un niveau individuel que sur l'évolution des schémas culturels et sociaux tels qu'ils sont représentés à travers différentes ressources.

Conclusion

La mémoire collective désigne un ensemble de représentations cruciales pour la définition de l'identité collective. Nous avons montré que ce concept permet de comprendre comment les hommes utilisent cette capacité cognitive lors de comportements sociaux et culturels, tant pour la formation de représentations communes (Mécanisme 1; Figure 1), la mise en récit de souvenirs représentatifs (Mécanisme 2; Figure 1) ou l'influence des ressources culturelles sur la cognition (Mécanisme 3; Figure 1). De nombreuses disciplines ont cherché à décrire les dynamiques qui la gouvernent, aboutissant à une opposition relativement artificielle entre un niveau d'explication purement social et un niveau individuel. La prise de conscience de ce dualisme est à la source du virage social et culturel des neurosciences et de façon déterminante un processus parallèle de changement paradigmatique, quoique encore limité, se retrouve également du côté des sciences humaines et sociales. Nous avons montré que le concept de schéma offre des pistes intéressantes pour comprendre davantage l'inscription des souvenirs collectifs et historiques dans le cerveau. La mémoire collective ne repose pas uniquement sur un socle de souvenirs communs mais sur une compréhension commune et

concertée des souvenirs. Cette compréhension prend sa source à la fois dans les structures représentationnelles préexistantes en mémoire, et dans l'effort exercé par le sujet pour relier et réduire les événements nouveaux à d'autres représentations, individuelles ou culturelles, d'événements passés.

Ces modèles ont des implications multiples, tant pour les neurosciences que pour les sciences humaines et sociales. Plutôt que de tirer des leçons de l'histoire, cette approche suggère qu'il est impossible pour une société et pour les individus qui la composent de comprendre l'évolution des événements présents sans se rattacher à une identité commune qui se fonde sur certains événements passés, sélectionnés et valorisés pour leur pertinence dans la mémoire collective. Ces évolutions ont également des répercussions sur le plan neuropsychologique et neuroscientifique dans la mesure où elles entraînent une transformation de leur objet d'étude, qui n'est plus simplement un système biologique isolé mais intégré et dépendant dans son évolution et son maintien d'interactions avec un environnement social et culturel. Elles engagent également une évolution des modèles utilisés pour expliquer l'intégration et la conservation de souvenirs à des composantes diverses, qui ne se réduisent pas à l'expérience propre de l'individu mais peuvent s'étendre aux représentations du groupe, de la catégorie sociale ou de l'histoire.

Les schémas permettent également d'introduire une cohérence supplémentaire entre les deux niveaux d'explication. Les concepts sociologiques se sont souvent développés sur la base d'une métaphore entre l'individu et la société, postulant par-là que certains phénomènes sont suffisamment semblables pour qu'ils puissent être décrits par analogie et par ressemblance. Le concept de schéma permet ce transfert direct entre les différentes analyses et pourrait ainsi devenir une articulation importante dans le dialogue entre descriptions psychologiques et sociales de la mémoire collective. Cette utilisation des schémas dépasse par ailleurs la description qu'en ont fournie jusqu'à présent les neurosciences cognitives et permet de formuler de nouvelles hypothèses qui pourront se soumettre à l'expérimentation. Les neurosciences pourraient définir comment les schémas sont susceptibles d'évoluer en fonction de nouvelles expériences mais également à travers la simple réflexion ou l'intégration de nouveaux contenus, notamment par des biais culturels. L'influence des ressources culturelles sur la mémoire individuelle reste peu étudiée, particulièrement sur le plan neural. Le développement de méthodes permettant d'étudier de façon conjointe les dynamiques neurales et l'évolution des représentations culturelles permettraient de mieux cerner les logiques d'imprégnation de la culture dans le cerveau en établissant également des liens originaux

entre sciences sociales et neurosciences. Enfin, le développement des bases neurales de la mémoire collective et de la dimension culturelle de la cognition peuvent conduire l'histoire et la philosophie à reconsidérer leurs modalités d'écriture en décryptant certaines logiques structurant récits et représentations symboliques à travers les mécanismes cognitifs qui en sont à l'origine.

Remerciements. Nicolas Legrand bénéficie d'une allocation doctorale de l'Observatoire B2V des Mémoires.

Références

- Anastasio, T.J., Ehrenberger, K.A., Watson, P., and Zhang, W. (2012). Individual and Collective Memory Consolidation: Analogous Process on Different Levels. MIT Press, Cambridge.
- Anderson, M.C., and Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410, 366–369.
- Anderson, M.C., Bjork, R.A., and Bjork, E.L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *J Exp Psychol Learn Mem Cogn*, 20, 1063–1087.
- Assmann, J. (2008). Communicative and Cultural Memory. In: A. Erll and A. Nünning eds., *Cultural Memory Studies: An International and Interdisciplinary Handbook*, de Gruyter, Berlin, pp. 253–262.
- Barash, J.A. (2006). Qu'est-ce que la mémoire collective? *Rev. Métaphysique Morale*, 50, 185.
- Barnier, A.J., and Sutton, J. (2008). From individual to collective memory: theoretical and empirical perspectives. *Memory*, 16, 177–182.
- Barnier, A.J., Sutton, J., Harris, C.B., Wilson, R.A., and Marsh, L. (2008). A conceptual and empirical framework for the social distribution of cognition: The case of memory. *Cogn Syst Res*, 9, 33–51.
- Bartlett, F. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, New York.
- Berlin, H.A. (2011). The Neural Basis of the Dynamic Unconscious. *Neuropsychanalysis*, 13, 5–71.
- Brown, N.R., Lee, P.J., Krslak, M., Conrad, F.G., Hansen G.B., Havelka, J., and Reddon, J.R. (2009). Living in history: How war, terrorism, and natural disaster affect the organization of autobiographical memory: Research Report. *Psychol Sci*, 20, 399–405.
- Brown, A.D., Kouri, N., and Hirst, W. (2012a). Memory's Malleability: Its Role in Shaping Collective Memory and Social Identity. *Front Psychol*, 3, 257.
- Brown, A.D., Kramer, M.E., Romano, T.A., and Hirst, W. (2012b). Forgetting Trauma: Socially Shared Retrieval-induced Forgetting and Post-traumatic Stress Disorder. *Appl Cogn Psychol*, 26, 24–34.
- Burman, J.T. (2012). History from within? Contextualizing the new neurohistory and seeking its methods. *Hist Psychol*, 15, 84–99.

- Cacioppo, J.T., and Decety, J. (2011). Social neuroscience: challenges and opportunities in the study of complex behavior. *Ann NY Acad Sci*, 1224, 162–173.
- Cacioppo, J.T., and Cacioppo, S. (2013). Social Neuroscience. *Perspect Psychol Sci*, 8, 667–669.
- Casey, E. (1987). Remembering: a phenomenological study. Indiana University Press, Bloomington.
- Chaumont, J.M. (2002). La concurrence des victimes: génocide, identité, reconnaissance. La Découverte, Paris.
- Cikara, M., and Van Bavel, J.J. (2014). The Neuroscience of Intergroup Relations: An Integrative Review. *Perspect Psychol Sci*, 9, 245–274.
- Coman, A., and Hirst, W. (2012). Cognition through a social network: the propagation of induced forgetting and practice effects. *J Exp Psychol Gen*, 141, 321–336.
- Coman, A., Brown, A.D., Koppel, J., and Hirst, W. (2009a). Collective memory from a psychological perspective. *Int J Polit Cult Soc*, 22, 125–141.
- Coman, A., Manier, D., and Hirst, W. (2009b). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of september 11 memories: Research article. *Psychol Sci*, 20, 627–633.
- Coman, A., Stone, C.B., Castano, E., and Hirst, W. (2014). Justifying Atrocities: The Effect of Moral-Disengagement Strategies on Socially Shared Retrieval-Induced Forgetting. *Psychol Sci*, 25, 1281–1285.
- Cuc, A., Ozuru, Y., Manier, D., and Hirst, W. (2006). On the formation of collective memories: the role of a dominant narrator. *Mem Cognit*, 34, 752–762.
- Cuc, A., Koppel, J., and Hirst, W. (2007). Silence is not golden: a case for socially shared retrieval-induced forgetting. *Psychol Sci*, 18, 727–733.
- Dessalles, J.-L. (2007). Why we talk – The evolutionary origins of language. Oxford University Press, New York.
- Dominguez Duque, J.F., Turner, R., Lewis, E.D., and Egan, G. (2010). Neuroanthropology: a humanistic science for the study of the culture-brain nexus. *Soc Cogn Affect Neurosci*, 5, 138–147.
- Dosse, F. (2010). Renaissance de l'événement. Un défi pour l'historien: entre sphinx et phénix. Presses universitaires de France, Paris.
- Dunsmoor, J.E., Murty, V.P., Davachi, L., and Phelps, E.A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, 520, 345–348.
- Durkheim, E. (1912). Les formes élémentaires de la vie religieuse : le système totémique en Australie. Presses universitaires de France, Paris, 1912, 2012.
- Echterhoff, G., Higgins, E.T., and Levine, J.M. (2009). Shared Reality: Experiencing Commonality With Others' Inner States About the World. *Perspect Psychol Sci*, 4, 496–521.
- Edelson, M., Sharot, T., Dolan, R.J., and Dudai, Y. (2011). Following the crowd: brain substrates of long-term memory conformity. *Science*, 333, 108–111.
- Eustache, F., Ganascia, J., Jaffard, R., Peschanski, D., Stiegler, B., and Sineux, P. (2014). Mémoire et oubli. le Pommier, Paris.
- Franks, D.D., and Turner, J.H. (2013). Handbook of Neurosociology. Springer Netherlands, Dordrecht.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci*, 11, 127–138.
- Gagnepain, P., Henson, R.N., and Anderson, M.C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proc Natl Acad Sci USA*, 111, E1310–E1319.
- Ghosh, V.E., and Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, 53, 104–114.
- Ghosh, V.E., Moscovitch, M., Melo Colella, B., and Gilboa, A. (2014). Schema Representation in Patients with Ventromedial PFC Lesions. *J Neurosci*, 34, 12057–12070.
- Glăveanu, V., and Yamamoto, K. (2012). Bridging history and social psychology: what, how and why. *Integr Psychol Behav Sci*, 46, 431–439.
- Halbwachs, M. (1925). Les cadres sociaux de la mémoire. Albin Michel, Paris, 1994, p.VI.
- Han, S., and Ma, Y. (2015). A Culture – Behavior – Brain Loop Model of Human Development. *Trends Cogn Sci* (sous presse).
- Harris, C.B., Barnier, A.J., Sutton, J., and Keil, P.G. (2014). Couples as socially distributed cognitive systems: Remembering in everyday social and material contexts. *Mem Stud*, 7, 285–297.
- Head, H., and Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, 34, 102–254.
- Henson, R.N., and Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20, 1315–1326.
- Hirst, W., and Manier, D. (2008). Towards a psychology of collective memory. *Memory*, 16, 183–200.
- Hirst, W., and Echterhoff, G. (2012). Remembering in Conversations: The Social Sharing and Reshaping of Memories. *Annu Rev Psychol*, 63, 55–79.
- Hirst, W., and Rajaram, S. (2014). Toward a social turn in memory: An introduction to a special issue on social memory. *J Appl Res Mem Cogn*, 3, 239–243.
- Jovchelovitch, S. (2012). Narrative, memory and social representations: a conversation between history and social psychology. *Integr Psychol Behav Sci*, 46, 440–456.
- Kim, H.S., and Sasaki, J.Y. (2014). Cultural neuroscience: biology of the mind in cultural contexts. *Annu Rev Psychol*, 65, 487–514.
- Knapp, A. (2014). Les Français sous les bombes alliées: 1940–1945. Tallandier, Paris.
- Knights, M. (2012). Taking a historical turn: possible points of connection between social psychology and history. *Integr Psychol Behav Sci*, 46, 584–598.
- Lavabre, M.-C. (1994). Le fil rouge: sociologie de la mémoire communiste. Presses de la Fondation nationale des sciences politiques, Paris.

- Lavabre, M.-C. (2013). De l'oubli dans les réflexions sur la mémoire « collective » ou « sociale. » In: *Mémoire et Mémorialisation*, D. Peschansky, ed. Hermann, Paris, pp. 11–24.
- Ledoux, S. (2012). Écrire une histoire du « devoir de mémoire ». *Débat*, 170, pp. 175–185.
- Liu, J.H., Paez, D., Hanke, K., Rosa, A., Hilton, D.J., Sibley, C.G., Cabecinhas, R., Zaromb, F., Garber, I.E., Leong, C.-H., Moloney, G., Valchev, V., Gastardo-Conaco, C., Huang, L.-L., Quek, A.-H., Techio, E., Sen, R., van Osch, Y., Muluk, H., Wagner, W., Wang, F., Khan, S. S., Licata, L., Klein, O., Laszlo, J., Fulop, M., Cheung, J. C.-K., Yue, X., Youssef, S. B., Kim, U., Park, Y., Puch-Bouwman, J., Hassall, K., Adair, J., Unik, L., Spini, D., Henchoz, K., Bohm, G., Selart, M., Erb, H.-P., Thoben, D. F., Leone, G., Mastrovito, T., Atsumi, T., Suwa, K.-I. (2012). Cross-Cultural Dimensions of Meaning in the Evaluation of Events in World History?: Perceptions of Historical Calamities and Progress in Cross-Cultural Data From Thirty Societies. *J Cross Cult Psychol*, 43, 251–272.
- Locke, J. (1689). *Essai sur l'entendement humain*. Vrin, Paris, 2001.
- Mandressi, R. (2011). Le temps profond et le temps perdu. *Rev Hist Sci Hum*, 25, 165.
- Manier, D., and Hirst, W. (2007). A cognitive taxonomy of collective memories. In: A. Erll and A. Nünning, eds. *Cultural Memory Studies: An International and Interdisciplinary Handbook*, de Gruyter, Berlin.
- Nietzsche, F. *Considérations inactuelles*. Gallimard, Paris, 1992.
- Nora, P. (1984). Entre mémoire et histoire: la problématique des lieux. In: *Les Lieux de Mémoire*. Tome 1: La République, P. Nora, ed. Gallimard, Paris.
- Olick, J.K. (1999). Collective Memory: The Two Cultures. *Sociol Theory*, 17, 333–348.
- Olick, J.K., and Robbins, J. (1998). Social Memory Studies: From “Collective Memory” to the Historical Sociology of Mnemonic Practices. *Annu Rev Sociol*, 24, 105–140.
- Pennebaker, J.W., and Gonzales, A. (2009). Making history: social and psychological process underlying collective memory. In: *Memory in Mind and Culture*, P. Boyer, and J. V. Wertsch, eds., Cambridge University Press, Cambridge, pp. 110–129.
- Pennebaker, J.W., Paez, D., and Deschamps, J.-C. (2006). The Social Psychology of History: Defining the Most Important Events of the Last 10, 100, and 1000 Years. *Psicol Política*, 32, 15–32.
- Peschanski, D. (2012). La mémoire dans l'histoire: régimes de mémorialité et conditions de la mise en récit mémoriel. In: *Les Années Noires, 1938-1944*. Hermann, Paris, pp. 387–401.
- Peschanski, D. (2013). *Mémoire et mémorialisation*. Hermann, Paris.
- Peschanski, D., and Maréchal, D. (2013). *Les chantiers de la mémoire*. INA, Bry-sur-Marne.
- Ricoeur, P. (2000). *La mémoire, l'histoire, l'oubli*. Ed. du Seuil, Paris.
- Roediger, H.L., and Abel, M. (2015). Collective memory: a new arena of cognitive study. *Trends Cogn Sci*, 19, 10–12.
- Roediger, H.L., Zaromb, F.M., and Butler, A.C. (2009). The role of repeated retrieval in shaping collective memory. In: *Memory in Mind and Culture*: P. Boyer, and J.V. Wertsch, eds. Cambridge University Press, Cambridge.
- Rouso, H. (1990). *Le syndrome de Vichy: de 1944 à nos jours*. Ed. du Seuil, Paris.
- Rubin, D.C., and Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychol Rev*, 122, 1–23.
- Schacter, D.L., Addis, D.R., and Buckner, R.L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci*, 8, 657–661.
- Smail, D.L. (2008). *On deep history and the brain*, University of California Press, Berkeley.
- Stone, C.B., and Hirst, W. (2014). (Induced) Forgetting to form a collective memory. *Mem Stud*, 7, 314–327.
- Stone, C.B., Barnier, A.J., Sutton, J., and Hirst, W. (2010). Building consensus about the past: schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory*, 18, 170–184.
- Stone, C.B., Coman, A., Brown, A.D., Koppel, J., and Hirst, W. (2012). Toward a Science of Silence: The Consequences of Leaving a Memory Unsaid. *Perspect Psychol Sci*, 7, 39–53.
- Stone, C.B., Barnier, A.J., Sutton, J., and Hirst, W. (2013). Forgetting our personal past: socially shared retrieval-induced forgetting of autobiographical memories. *J Exp Psychol Gen*, 142, 1084–1099.
- Stone, C.B., van der Haegen, A., Luminet, O., and Hirst, W. (2014). Personally relevant vs. nationally relevant memories: An intergenerational examination of World War II memories across and within Belgian French-speaking families. *J Appl Res Mem Cogn*, 3, 280–286.
- Sutton, J., Animates, H., and Systems, D. (2008). Between Individual and Collective Memory: Coordination, Interaction. *Soc Res*, (New York) 75, 23–48.
- Tavani, J.L., Collange, J., Rateau, P., Rouquette, M.-L., and Sanitioso, B.R. (2015). Tell me what you remember and I will know who you are: The link between collective memory and social categorization. *Gr Process Intergr Relations*, sous presse.
- Tronson, N.C., and Taylor, J.R. (2007). Molecular mechanisms of memory reconsolidation. *Nat Rev Neurosci*, 8, 262–275.
- Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., and Morris, R.G.M. (2007). Schemas and memory consolidation. *Science*, 316, 76–82.
- Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., and Morris, R.G.M. (2011). Schema-Dependent Gene Activation. *Science*, 333, 891–895.

- Van Bavel, J.J., and Cunningham, W.A. (2012). A Social Identity Approach to Person Memory: Group Membership, Collective Identification, and Social Role Shape Attention and Memory. *Personal Soc Psychol Bull*, 38, 1566–1578.
- van Kesteren, M.T.R., Fernández, G., Norris, D.G., and Hermans, E.J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proc Natl Acad Sci USA*, 107, 7550–7555.
- van Kesteren, M.T.R., Ruiter, D.J., Fernández, G., and Henson, R.N. (2012). How schema and novelty augment memory formation. *Trends Neurosci*, 35, 211–219.
- van Kesteren, M.T.R., Beul, S.F., Takashima, A., Henson, R.N., Ruiter, D.J., and Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. *Neuropsychologia*, 51, 2352–2359.
- Viard, A., Chételat, G., Lebreton, K., Desgranges, B., Landeau, B., de La Sayette, V., Eustache, F., and Piolino, P. (2011). Mental time travel into the past and the future in healthy aged adults: An fMRI study. *Brain Cogn*, 75, 1–9.
- Wagoner, B. (2012). Culture in Constructive Remembering. In: *The Oxford Handbook of Culture and Psychology*: J. Vasnier, ed., Oxford University Press, Oxford, UK, pp. 1034–1055.
- Wagoner, B. (2013a). Bartlett's concept of schema in reconstruction. *Theory Psychol*, 23, 553–575.
- Wagoner, B. (2013b). Culture and mind in reconstruction: Bartlett's analogy between individual and group processes. In: *Doing Psychology under New Conditions*: A. Marvakis, J. Motzkau, D. Painter, R. Ruto-Korir, G. Sullivan, S. Triliva, and M. Wiesser, eds., Captus Press Concord, ON, pp. 273–278.
- Wertsch, J. V. (2008). The Narrative Organization of Collective Memory. *Ethos*, 36, 120–135.
- Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., and Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *J Neurosci*, 31, 1563–1569.
- Zerubavel, E. (1996). Social memories: Steps to a sociology of the past. *Qual Sociol*, 19, 283–299.

Sleep differentially improves event-based prospective memory in young and old adults: a virtual reality study

Nicolas Legrand¹, Stéphane Rehel¹, Gregory Lecouvey¹, Alice Laniepce¹,
Françoise Bertran^{1,2}, Philippe Fleury³, Sophie Madeleine³, Francis Eustache¹,
Béatrice Desgranges¹, and Géraldine Rauchs¹

¹Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France

²Unité d'exploration et de traitement des troubles du sommeil, University Hospital, Caen, France

³Centre Interdisciplinaire de Réalité Virtuelle, UNICAEN, 14000 Caen, France

Correspondance:

Dr Géraldine Rauchs,

Inserm-EPHE-UNICAEN U1077, GIP Cyceron, Bd Becquerel

BP 5229, 14074 CAEN cedex 5, France

Tel: 02.31.47.01.34

E-mail: *geraldine.rauchs@inserm.fr*

Abstract

Prospective memory (PM) refers to our ability to execute delayed intentions at the appropriate moment. Studies have highlighted the role of sleep in the consolidation of PM intentions, but whether this benefit is preserved during aging is still debated. In this study, we used a virtual environment to assess the consolidation of intentions after 12h filled either with daytime wakefulness or with nighttime sleep in 31 young and 29 older healthy volunteers. We manipulated the level of monitoring (i.e. attentional engagement required to perform the action) and congruence of the intentions. Older adults showed a reduced PM consolidation after sleep as compared to young ones. Overnight improvement in PM performance was positively correlated with subjective sleep efficiency in young but not in older adults. When assessing the effect of monitoring and congruency, we observed a pattern of results suggesting that the two populations may rely on different strategies: young adults favoring sleep-mediated plasticity to rapidly acquire new unfamiliar information, while older adults may rather rely on a knowledge-based strategy, highly effective but less prone to changes.

Keywords: sleep, prospective memory, aging, virtual reality, memory schema, plasticity

Introduction

Sleep favors consolidation of recently acquired information into long term stores (Diekelmann and Born, 2010; Rauchs et al., 2005). This effect has been widely reported for memories of past events, but other studies also suggest that sleep may preferentially strengthen memories relevant for future behaviors, if they are either required for a later memory test (Rauchs et al., 2011; van Dongen et al., 2012; Wilhelm et al., 2011) or associated with an anticipated reward (Fischer and Born, 2009). Prospective memory (PM), which refers to the ability to remember to execute an intention in the future (McDaniel, 2007), is typically representative of such a future-oriented memory process. In the last few years, studies documented that sleep also benefits PM performance (Scullin and McDaniel, 2010), an effect mainly attributable to Slow-Wave Sleep (Diekelmann et al., 2013) and mediated by a facilitation of spontaneous-associative retrieval processes during PM recall (Diekelmann et al., 2013).

Two components are essential to correctly execute delayed intentions: a prospective component and a retrospective one (Einstein and McDaniel, 1990). The prospective component involves remembering that something has to be done at the appropriate moment. The retrospective component refers to the content of the intention, in other words, remembering what has to be done. Retrieval of an intention can be triggered either by the occurrence of an external event, namely the prospective cue (e.g. “take the cake out of the oven when the timer rings”), or after a defined amount of time (e.g. “take the cake out of the oven after 30 minutes”). These two forms of PM respectively refer to *event-based* PM (EBPM) and *time-based* PM (TBPM) (Einstein and McDaniel, 1990). However, while intentions are held in mind, participants are engaged in other activities (named the “*ongoing task*”), distinguishing PM from simple planning. Thus, retrieving the prospective component may engage different level of monitoring processes as the detection of the prospective cue can be either automatic as the participant interact with the environment, or require attentional processes if an

active research is needed.

Once the prospective component has been correctly retrieved, participants engage memory retrieval processes to recall what has to be done (i.e., the retrospective component). These processes may be influenced by the strength of the link between the prospective and retrospective components. The latter can be semantically linked to the former (e.g. “*At the cafeteria, buy a coffee*”), i.e. congruent, or not (e.g. “*At the cafeteria, rent a DVD*”). Whether the benefit of sleep on PM is modulated by the level of monitoring and congruence of intentions still remains an open question.

PM has generally been found impaired in aging (Einstein and McDaniel, 1990; Gonneaud et al., 2011; McDaniel and Einstein, 2011; Uttl, 2008; Lecouvey et al., 2017), although this effect has also been described as paradoxical. Older adults can indeed show similar or higher performance as compared to young individuals when the PM task relies on automatic processes, whereas PM remembering implying controlled processes is generally impaired. Moreover, this effect can be influenced by the ecological dimension and the familiarity of the participant with the task and the environment (Aberle et al., 2010; Schnitzspahn et al., 2010).

Sleep-dependent consolidation of “classical” episodic memories has been investigated in older participants (Mander et al., 2017), providing discrepant results. Some studies showed that this process is impaired (Backhaus et al., 2007; Cherdieu et al., 2014; Mander et al., 2013) and may be explained by well-described changes in sleep architecture, especially the dramatic decrease in slow-wave sleep (SWS). In contrast, other studies reported that this process is preserved (Aly and Moscovitch, 2010; Wilson et al., 2012), especially when the material to learn is personally relevant (Aly and Moscovitch, 2010). Recently, (Scullin et al., 2017) reported that PM consolidation during a nap is impaired in older adults, and that this effect would be mediated by a decrease in REM sleep duration. However, we still have a poor understanding of the cognitive basis of this impairment, especially if it relies on memory or attentional components engaged during retrieval.

In this context, our study aimed at investigating the effect of retention intervals filled with sleep

or daytime wakefulness on the recall of *event-based* intentions in young and older participants. We designed a PM task implemented in a complex virtual environment aiming at optimizing the ecological aspects of PM. To assess the effect of sleep on different retrieval processes for EBPM, we also manipulated two variables: 1) the level of monitoring processes and 2) the strength of the semantic link between the prospective and the corresponding retrospective component to determine whether congruent intentions are better remembered than incongruent ones.

Materials and methods

Participants

Thirty-one young university students (15 males, mean age \pm SD: 22 ± 2.9 y, range 18-34) and 29 older adults (14 males, mean age \pm SD: 62.6 ± 4.5 y, range 55-72) participated in this study which was approved by the ethics committee (CPP Nord Ouest III). Written informed consent was obtained from each participant after a detailed presentation of the study. Participants' characteristics are presented in **Table 1**.

All participants were native French speakers, right handed, had normal or corrected to normal vision and good health as revealed by a medical examination. During the study, participants were not on medication, except for oral contraceptive. Among the 60 participants, 56 had a Pittsburgh Sleep Quality Index (PSQI) (Buysse et al., 1989) below or equal to 5, indicating no sleep disturbances. Four participants had PSQI scores of 6 or 7, but an interview with a sleep expert neurologist confirmed the absence of serious sleep difficulties. None of the participants had worked night shifts during the preceding year. All participants had normal scores on the 21-item Beck Depression Inventory (Beck, 1961) (score [?] 8). Extreme chronotypes were excluded according to the morningness-eveningness questionnaire (Horne and Ostberg, 1976)(scores [?] 32 or [?] 70). The absence of signs of cognitive decline in the group of Older adults was confirmed by the Mini

Mental State Examination (Folstein et al., 1975) (scores > 27). All volunteers followed a constant sleep schedule (according to their own sleep-wake schedule +/- 1 hour) throughout the course of the study (4 weeks from the medical examination to the last testing session). Participants received a financial compensation for their participation.

Prospective memory task

Virtual environment

PM was assessed using a complex virtual environment originally provided by the CIREVE (Interdisciplinary Centre of Virtual Reality, UNICAEN, France) and developed with Unity 4.3 software (www.unity3d.com/). This environment recreated the ground floor of the museum “*Memorial*” located in Caen and dedicated to World War II. The virtual museum was displayed using a beamer on a 244 cm widescreen. Each subject was tested individually, comfortably installed at approximately 400 cm of the widescreen. Any faintness possibly induced by virtual reality or stress due to the experiment was subsequently reported by the participants. Subjects interacted with the virtual environment from a first person perspective using a USB wired controller with 2 joysticks and 3 buttons. One joystick allowed to move in the environment, and the other controlled head movements and visual field orientation. The three buttons allowed participants to display a clock, a map of the environment and a list of events (see below).

Fifteen black and white pictures, representing several striking events of the Second World War and present in the original museum, were displayed on the walls of the virtual environment. As an ongoing task, subjects had to look at, and then remember, a set of fifteen different pictures for each session (10 minutes, WAKE, and SLEEP conditions, see below Experimental Design and Procedure). A recognition test was proposed after the visit of the virtual museum. For the purpose of this study, a clock, a map and a menu displaying descriptions of events occurring in the museum

were added. These elements were accessible during exploration via the 3 buttons of the controller. The clock displayed a virtual time. The map represented the virtual environment with the name of relevant locations and the subject's position. Finally, the menu consisted of a list of six sentences, each describing an event currently occurring in the virtual museum. Each event went through three different states during the test (e.g., state 1: "*the checkouts are closed*"; state 2: "*the checkouts will soon open*" ; state 3: "*the checkouts are open*") in an unpredictable timing. The third state was the prospective cue of an intention that participant had to detect for four among those six events (two were foils). Subjects had then to regularly monitor the status of those events.

Intentions

For each session, the PM task consisted in the encoding and recall of 8 EBPM intentions. Among those intentions, 4 implied a high level of monitoring and required to check regularly the events menu to detect the corresponding prospective cues, and 4 required a low level of monitoring because the prospective cue was a particular location in the virtual museum visited by the participants. Among both high and low monitoring intentions, 2 were semantically congruent (link-EBPM) and presented an evident link between the prospective cue and the retrospective component (e.g. low monitoring condition: "*At the cafeteria, order a coffee*", high monitoring condition: "*When the ticket office opens, buy an entrance ticket*") whereas for the 2 others, it was not the case (noLink-EBPM; e.g. low monitoring condition: "*At the cafeteria, rent a DVD*"; high monitoring condition: "*When your friend leaves the museum, take a picture of the map of the world*").

Intentions differed between the three sessions (**10 minutes** , **WAKE** and **SLEEP**). Each intention was described through a unique sentence. The first part of the sentence contained the prospective component of the intention (e.g. "*At the cafeteria . . .*") and the last part referred to the retrospective component, the action that participants intended to execute (e.g. "*. . . order a black coffee*").

Experimental design and procedure

This study comprised three sessions. Before the first session, we proposed a familiarization phase during which participants visited the entire virtual museum and exercised to use the controller. After this familiarization period, participants were informed that they will again be immersed in the virtual museum and they will have to execute a series of actions. We also explained that some of the actions will have to be performed when they will be in a particular location of the virtual environment (Low Monitoring EBPM) and others when a particular event will occur in the virtual museum (High Monitoring EBPM).

In the first session, participants learned a series of 8 intentions. Retrieval was assessed after a 10-min delay (**10 min condition**). One group of subject (15 young and 15 older adults) performed this session at 8 a.m. (**MORNING condition**) and the other (16 young and 14 older adult) at 8 p.m. (**EVENING condition**). This first session was used to control for a potential circadian confound as learning and recall of intentions occur at different times of the day for the SLEEP and WAKE conditions (see **Figure S1**).

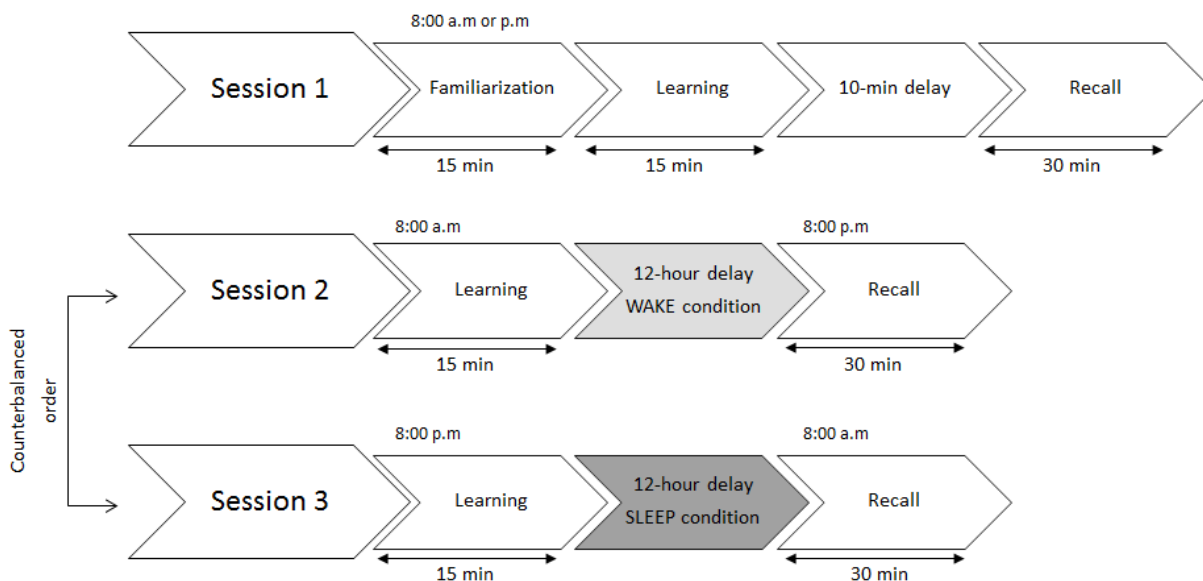


Figure 1: Experimental design.

In the second and the third sessions, recall of intentions was probed after a 12-hour delay filled with either daytime wakefulness (**WAKE condition**) or nighttime sleep (**SLEEP condition**). In the **WAKE condition**, subjects encoded the 8 intentions at 8 a.m. and recalled them at 8 p.m (see **Figure S1**). In the **SLEEP condition**, intentions were encoded at 8 p.m. and recalled the next morning at 8 a.m. Volunteers participated in the three sessions in a within-subject design, the order of the Sleep and Wake conditions being counterbalanced across participants. All the sessions were performed at the University Hospital in Caen, with at least one week between sessions to minimize the risk of interference.

Encoding of intentions

For each session, the 8 intentions to learn were presented on the screen of a laptop. To ensure that all intentions were correctly encoded, a cued recall test was proposed after learning, presenting

the prospective component of the intentions and asking to recall the retrospective one. A feedback was given after each response, and the procedure repeated until subjects correctly recalled every intention two consecutive times.

Storage of intentions

After their arrival to the laboratory, participants completed the Karolinska Sleepiness Scale (KSS) (Akerstedt and Gillberg, 1990) to control the alertness on each session. In the first session, the 10-min interval was filled by completion of the Mill Hill vocabulary questionnaire (Raven, 1965) and the STAI A and B (Spielberger et al., 1983) measuring trait and state anxiety.

For the second and the third sessions, participants left the laboratory after learning the 8 intentions and came back for the retrieval phase 12 hours later. This period was covered either by regular sleep at home in the SLEEP condition, or by an equivalent period of daytime wakefulness during which participants were engaged in their usual activities.

Compliance to the sleep-wake schedule and the absence of naps were assessed using wrist actigraphy (Actiwatch, Cambridge Neurotechnology Ltd). Furthermore, upon their return to the laboratory for the retrieval phase, participants completed again the KSS and the St Mary's Hospital Sleep Questionnaire (Ellis et al., 1981) to assess sleep quality during the previous night. This questionnaire contains a subjective evaluation of sleep onset, awakening and the number of hours in bed, allowing to compute a subjective measure of sleep efficiency (subjective time spent asleep / subjective time spent in bed).

Retrieval of intentions

Once the 10-min or 12-hour delay had elapsed, subjects were again immersed in the virtual museum. They were reminded that they had to execute the learned intentions at the appropriate moment. To do so, they had to recall as accurately as possible the sentence describing the intention.

Each session lasted at least 15 minutes and participants were allowed to stay at most 20 minutes.

A total of six points were given for the correct recall of an intention. A maximum of two points were attributed for the correct recall of the prospective component (*High Monitoring EBPM intentions* : 2 points if the difference between the time at which the intention was executed and the time at which it was supposed to do it did not exceed 1 minute, 1 point if it did not exceed 2 minutes and 0.5 points if the difference was greater than 2 minutes; *Low Monitoring EBPM* : 2 points if recall occurred when entering for the first time in the place targeting retrieval, 1 point if it occurred after the second passing, and 0.5 point for 3 or more passings). As well, a maximum of two points were given for a correct recall of the retrospective component (2 points for the exact action, 1 point for a closely related action and 0.5 point for a slightly related action).

Statistical analyses

Statistical analyses consisted in factorial analyses of mixed models using the *aov_ez* function implemented in the *afex* package (Singmann et al., 2017) in R (version 3.4.1). Post hoc tests were carried out using the *lsmeans* package (Lenth, 2016) with the Tukey correction for pairwise comparison.

Results

Participants' characteristics are reported in Table 1. No differences were found in terms of educational level ($p=0.75$), anxiety level (STAI-B; $p=0.87$), Beck Depression Inventory ($p=0.7$) or PSQI score ($p=0.85$). We did not observed any significant difference of sleepiness as measured with the Karolinska Sleepiness Scale either between Groups ($F_{(1, 56)} = 0.01$, $\eta_g^2 > 0.0001$, $p=0.94$), or testing time ($F_{(1, 56)} = 0.85$, $\eta_g^2 = 0.01$, $p=0.36$) in the 10 minute condition. Similar results were reported in the delayed condition for Group ($F_{(1, 55)} = 0.00$, $\eta_g^2 > 0.0001$, $p=0.95$), Condition ($F_{(1, 55)}$)

= 2.39, $\eta_g^2=0.005$, $p=0.13$) or testing time ($F_{(1, 55)} = 0.45$, $\eta_g^2=0.001$, $p=0.51$). A visual inspection of actigraphy recordings confirmed schedule compliance and the absence of naps, for each subject, in the Sleep and Wake conditions respectively.

The analysis of behavioral data first consisted in ensuring that the difference in PM performance between Sleep and Wake conditions was not imputable to the time of day at which subjects learned the intentions and were tested (8 a.m. or 8 p.m.). To do so, we first analyzed the global PM performance (sum of scores for the prospective and retrospective components) with a two-way between-subjects ANOVA with both Age (Young vs Old) and time of day (8 a.m. vs 8 p.m.) as between subject factors. This analysis revealed no significant effect of age ($F_{(1, 56)} = 0.05$, $\eta_g^2=0.001$, $p=0.82$), time of day ($F_{(1, 56)} = 1.08$, $\eta_g^2=0.02$, $p=0.3$) nor interaction between these two factors ($F_{(1, 56)} = 0.01$, $\eta_g^2<0.001$, $p=0.91$), indicating that global PM performance was similar between Young and Older participants whenever learning and testing occurred (8 a.m. vs 8 p.m.).

To assess the improvement in global recall of event-based intentions due to sleep, we performed a mixed analysis of variance with Condition (WAKE, SLEEP) as within subject factor and Age (Young, Old) as between subject factor. This analysis revealed no significant effect of Age ($F_{(1, 58)} = 0.00$, $\eta_g^2=0.0001$, $p> 0.99$) but a significant effect of Condition ($F_{(1, 58)} = 63.75$, $\eta_g^2=0.21$, $p< 0.001$) and a significant Age x Condition interaction ($F_{(1, 58)} = 4.24$, $\eta_g^2=0.02$, $p=0.04$). Follow-up t-test on the difference SLEEP-WAKE for each group revealed that Young adult benefited globally more from sleep than Old ones ($t=2.0$, $p =0.04$; **Figure S2**, panel A).

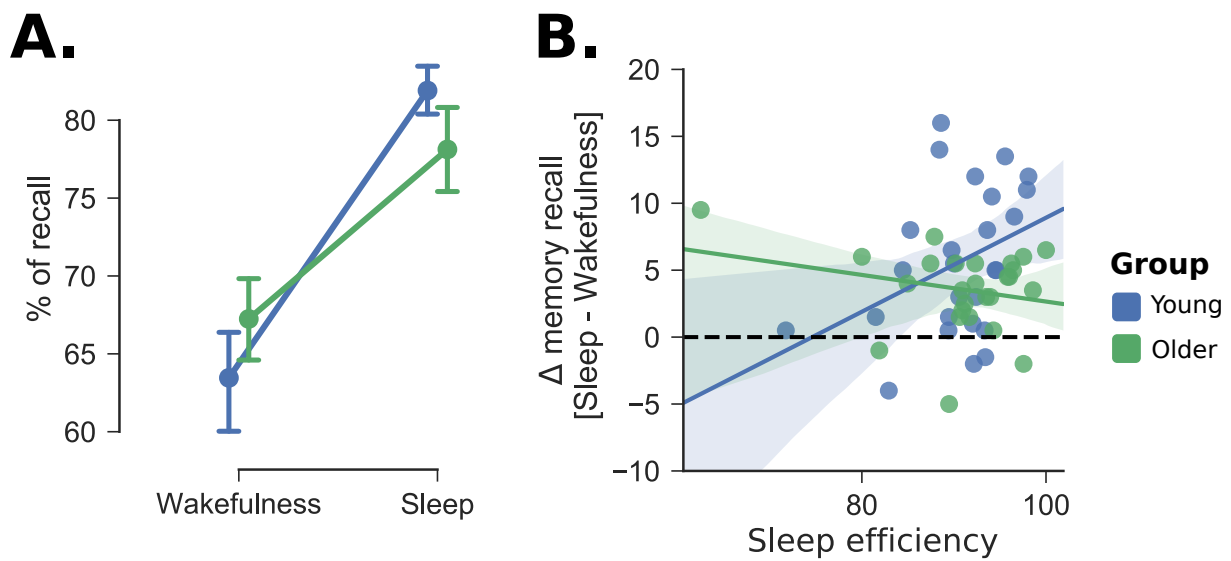


Figure 2: Effect of sleep on event-based PM consolidation. **A.** Recall of the PM intentions after 12h filled with sleep or wakefulness. **B.** Correlation between improvement of PM recall and sleep efficiency for Young and Older adults.

We then searched for correlations between the PM improvement (defined as follows: score in the SLEEP condition – score in the WAKE condition) and subjective sleep efficiency computed with information gathered from the St Mary’s Hospital Sleep Questionnaire concerning the postlearning night. Data for sleep efficiency were not exploitable for 5 Young and 3 Old adults, so this analysis was performed only on 26 Old and 26 Young participants. As we hypothesized that sleep efficiency would be positively correlated to PM improvement, we used a one-tailed Spearman correlation test ($\rho > 0$). This analysis revealed a positive correlation in the group of Young adults ($\rho = 0.34$, $p = 0.04$) but not in Older participants ($\rho = -0.03$, $p = 0.57$). Results are illustrated in **Figure 2**, panel **B**.

Then, we focused our analyses on the influence of level of monitoring and congruency on sleep-dependent memory consolidation. We expected monitoring to influence the recall of the prospec-

tive component and congruency to influence the recall of the retrospective component. For each comparison, we first controlled for the presence of the effect in the 10-min condition. The analysis of the recall of the prospective component of low and high-monitoring intentions after 10 minutes was performed with a mixed ANOVA with the monitoring level (Low, High) as within-subject factor and Age (Young, Old) as between subject factor. This analysis revealed no significant effect of Age ($F_{(1, 58)} = 0.21, \eta_g^2=0.002, p =0.65$), but a significant effect of monitoring ($F_{(1, 58)} = 34.13, \eta_g^2=0.19, p<0.001$) and an interaction between these two factors ($F_{(1, 58)} = 6.52, \eta_g^2=0.04, p=0.01$). Post-hoc comparisons revealed that Young adults had better recall performance than Older participants for high-monitoring intentions ($p=0.048$). In contrast, there was no difference between groups for low-monitoring intentions ($p=0.20$; **Fig. 3, panel A3**). Then, we explored the effect of sleep and wakefulness on the recall of the prospective component of low-monitoring intentions. To do so, a mixed ANOVA with Condition (WAKE, SLEEP) as within subject factor and Age (Young, Old) as between subject factor was conducted. This analysis revealed a significant effect of Age ($F_{(1, 58)} = 4.67, \eta_g^2=0.05, p=0.03$), a significant effect of Condition ($F_{(1, 58)} = 24.58, \eta_g^2=0.11, p<0.001$) and a significant interaction between these two factors ($F_{(1, 58)} = 4.10, \eta_g^2=0.02, p=0.05$). Post-hoc comparisons revealed a difference between Young and Older adults in the WAKE condition ($p< 0.01$) but not in the SLEEP one ($p=0.51$; **Figure 3, panel B**).

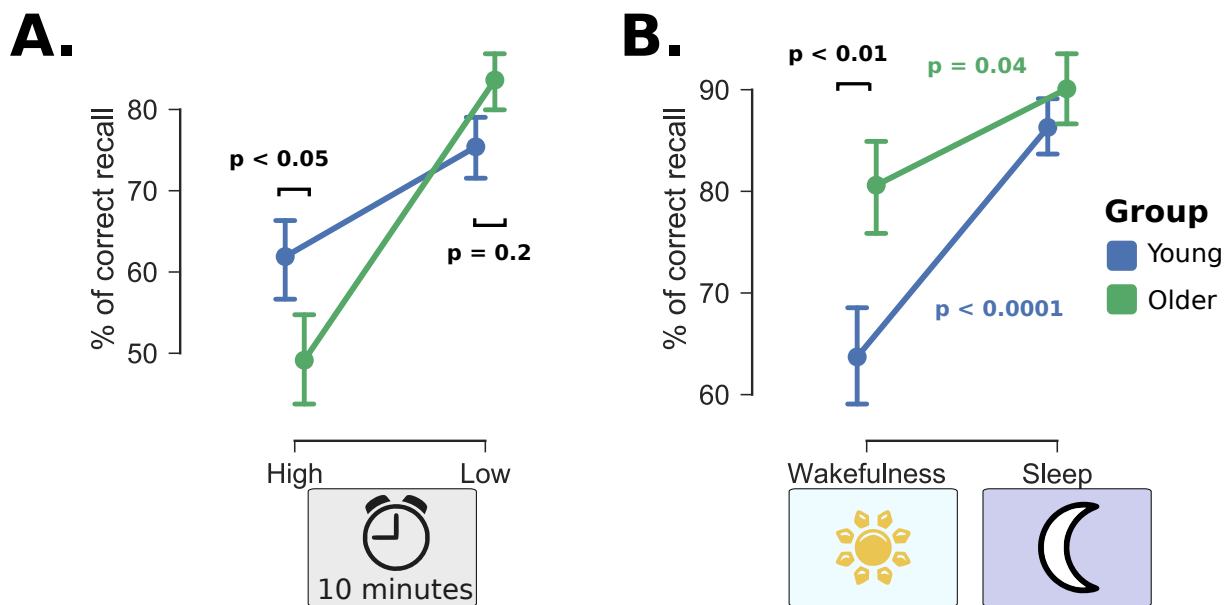


Figure 3: Effect of sleep on PM consolidation for intentions requiring a low level of attentional monitoring. **A.** Recall of the prospective component for intentions requiring High or Low attentional monitoring after 10 minutes. **B.** Recall of the prospective component for Low monitoring intentions after 12 hours filled with sleep or wakefulness.

Then, we examined the recall of the retrospective component of congruent and incongruent intentions after 10 minutes with a mixed ANOVA with Congruence (Congruent, Incongruent) as within subject factor and Age (Young, Old) as between subject factor. This analysis revealed a significant effect of Congruence ($F_{(1, 58)} = 5.82, \eta_g^2=0.04, p=0.02$) but no effect of Age ($F_{(1, 58)} = 0.38, \eta_g^2=0.004, p=0.54$) or interaction between the two factors ($F_{(1, 58)} = 0.04, \eta_g^2=0.0003, p=0.84$) indicating that, in both groups, participants performed equally when recalling the retrospective component of congruent and incongruent intentions (see **Figure S4**, panel A).

Focusing on the recall of incongruent intentions, we analyzed the sleep-dependent memory consolidation with a mixed ANOVA with Condition (WAKE, SLEEP) as within subject factor and Age

(Young, Old) as between subject factor. This analysis revealed a significant effect of Condition ($F_{(1, 58)} = 13.69, \eta_g^2=0.07, p<0.005$) but no effect of Age ($F_{(1, 58)} = 0.43, \eta_g^2=0.005, p =0.51$) nor interaction between these two factors ($F_{(1, 58)} = 2.25, \eta_g^2=0.01, p=0.14$). However, post-hoc revealed that young adult significantly improved overnight their performance for incongruent intentions ($p<0.001$), while older adults did not ($p=0.13$; **Figure 4**, panel **B**).

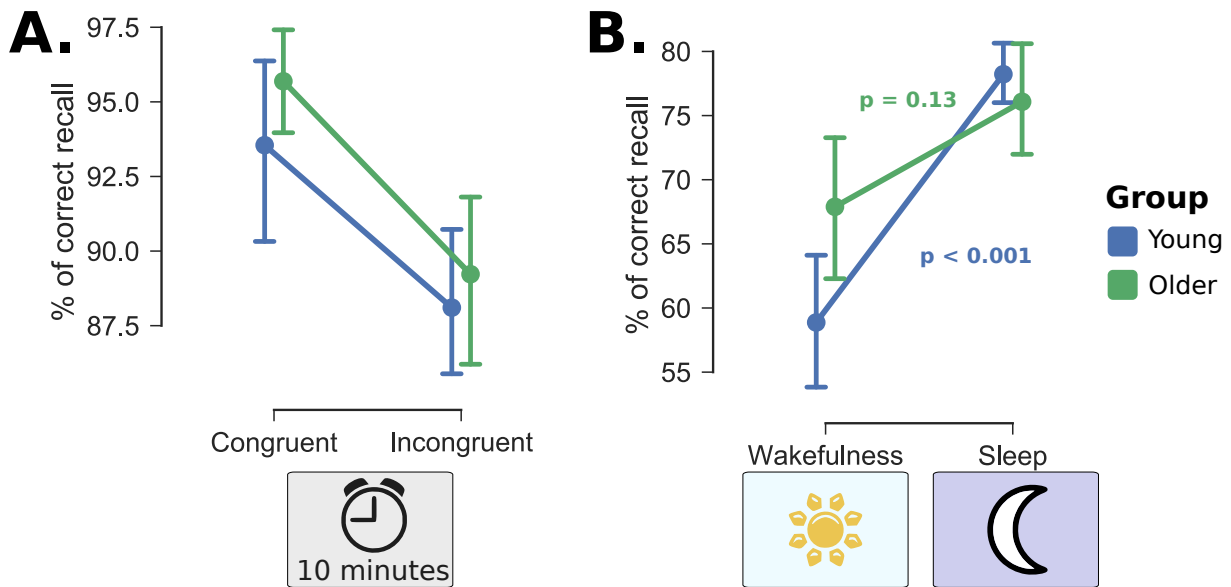


Figure 4: Effect of congruence on sleep-dependent PM consolidation. **A.** Recall of the retrospective component for congruent and incongruent intentions after a delay of 10 minutes. **B.** Recall of the retrospective component for incongruent intentions after 12 hours filled with Sleep or Wakefulness.

Discussion

This study aimed at investigating the effect of retention intervals filled with sleep or daytime wakefulness, on the recall of intentions in young and older participants. A short-term condition (recall

after 10 minutes) was added to control for potential circadian effects as participants did not learn and were not tested at the same time of day in the Sleep and Wake conditions. The analysis of this 10-min condition revealed that global PM performance was similar between young and older participants, whatever the time of testing. This result allowed to ensure that baseline performance is not influenced by possible difference in the level of alertness due to the time of the day. It also indicates that global PM performance, as assessed using a complex and realistic virtual environment is not impaired in older adults, provided that the delay is short.

When we contrasted performance after 12h retention intervals filled with daytime wakefulness or sleep, we found a significant influence of age on the sleep-dependent improvement of PM recall, which is in line with previously reported results both with retrospective ([Cherdiou et al., 2014](#); [Backhaus et al., 2007](#)) and prospective memory ([Scullin et al., 2017](#)). In addition, we observed a significant correlation between PM improvement and subjective sleep efficiency, but this effect was only found in Young participants. These results suggest that the PM impairment we report may be more directly related to sleep efficiency, while this conclusion would require additional investigation because it was also reported that older adults may have difficulties to correctly assess their sleep ([Silva et al., 2007](#); [Van et al., 2008](#)), thus biasing the result.

A more detailed analysis of the recall of the prospective component revealed that age and the level of monitoring interacted in the first condition comprising only 10 min between learning and recall. Compared to Young individuals, Older adults showed better performance for intentions requiring a low monitoring level, but lower performance for high-monitoring intentions. This result is consistent with previous observations suggesting greater difficulties for older adults when performing tasks requiring attentional resources and cognitive flexibility engaging the prefrontal cortex ([Burgess et al., 2001](#)), but similar or higher performance as compared to young adults under low-demanding or ecologically familiar tasks.

Interestingly, the difference between Young and Older adults for the recall of the prospective com-

ponent of low-monitoring intentions was also observable in the WAKE condition. However, after a night of sleep, Young adults significantly enhanced their performance and reached a level similar to Older participants. The recall of the prospective component under low-monitoring condition is thought to be mediated by spontaneous associative retrieval triggered when participants are confronted with the specific PM cue (i.e. a location in the virtual environment). When the interval between intention formation and retrieval is long, participants are thought to favor associative retrieval processes over monitoring (Scullin et al., 2013). As those processes are mainly dependent on the hippocampal formation (Gordon et al., 2011), we can hypothesize that they are more likely to be improved after a night of sleep as NREM sleep has been shown to favor the consolidation of hippocampal-dependent memories (Diekelmann and Born, 2010) and is determinant for PM consolidation (Diekelmann et al., 2013). Previous works reported such benefits of sleep on associative retrieval processes, inducing consolidation of intention in associative networks allowing subjects to recall intentions without costs on the ongoing task (Scullin and McDaniel, 2010). (Diekelmann et al., 2013) showed that subjects who did not slept after learning had impaired performance in the detection of relevant cues in an ongoing task. Our data indicate that Older adults tend to perform correctly on the recall of low-monitoring intentions without needing further consolidation during sleep. In contrast, Young adults, despite difficulties in the 10-minute and WAKE conditions, benefited from sleep. We suggest that young and older participants may rely on different strategies to perform the PM task. Young adults may rely on a plasticity-based process allowing greater improvement in performance if sufficient rest is provided, while Older adults may rather rely on their acquired knowledge, allowing greater performance on the first trial but with minor overnight improvement.

Concerning the recall of retrospective component of incongruent intentions, we found that young adults significantly increased their performance after a night of sleep while older showed a moderate change in performance. Retrieval of the retrospective component of an intention is thought

to engage memory retrieval processes that could be affected by the strength of the semantic link between the prospective and retrospective components. A previous study conducted in our laboratory reported that young adults have better recall performance of the retrospective component for coherent than for incoherent intentions (Gonneaud et al., 2012). Here, our result suggest moreover that age may modulate the consolidation process of incongruent intention.

Encoding and recalling information can be facilitated by preexisting related knowledges (Gilboa and Marlatte, 2017; Tse et al., 2007; van Kesteren et al., 2012). Memory schemas refer to the cognitive structures encoding gist representation of repeatedly encountered events (van Kesteren et al., 2012). The repetition of similar episodes (e.g. repeatedly buying a coffee in a cafeteria) can participate to the formation of an associative framework connecting core elements of those experiences. Those representations are thought to favor retrieval and encoding via the preferential engagement of the medial prefrontal cortex instead of the hippocampus (van Kesteren et al., 2014; van Kesteren et al., 2013; Tse et al., 2007) when encoded information relates to preexisting ones. This process is likely to be mediated by sleep-dependent memory consolidation (Landmann et al., 2014; Lewis and Durrant, 2011), both during SWS (Hennies et al., 2016) and REM sleep (Durrant et al., 2015), where it favors the oscillatory dynamic of memory reactivations (Groch et al., 2017).

But it was also supposed that this effect might be worsen by the incongruency of the learned intentions. Incongruent intentions still relate to preexisting knowledge as they do not contain new and neutral content for the participant. But their unusual association implies that the influence of memory may be modulated to prevent interferences. Consolidating the intention “*Rent a DVD at the cafeteria*”, will then engage a modulation of the preexisting memory framework automatically connecting “*cafeteria*” with “*buy a coffee*”. However, as aging is associated with both changes in sleep architecture (especially a decrease in SWS efficiency) and mPFC atrophy (Mander et al., 2013; Helfrich et al., 2018), we surmise that, in older participants, this plasticity allowing

the integration of unexpected or unusual information into memory frameworks would be more difficult.

The results reported in this study also raise the question of the potential effect of cognitive reserve on sleep and memory processes. Indeed, older participants included in this study were highly educated, education being a common proxy of cognitive reserve, and involved in many activities (Stern, 2012). We can surmise that individuals with high cognitive reserve may be more able to compensate for neural dysfunction due to disturbed sleep and may maintain good memory performance. However, such a compensation may be more efficient for information coherent with pre-existing knowledge, “unusual associations” requiring more efforts and resources to be consolidated.

Finally, one limitation of our study is the lack of insight into objective sleep quality and the electrophysiological activity underpinning memory consolidation as it could be measured with polysomnography. Futures studies using sleep recordings would be valuable to decipher the exact physiological contribution of PM enhancement and to contrast sleep profile of young adults as compared to older ones or pathological populations.

Conclusion

This study reports that PM consolidation during sleep is impaired among older adults, and that this impairment is correlated with subjective sleep efficiency. We replicated previous works evidencing the role of sleep in the integration of prospective components into associative frameworks as well as an influence of the congruency level for the recall of the retrospective component. The results obtained also suggest that young and older adults may rely on different strategies to perform the PM task. Young adults benefited more of an interval filled with sleep and compensated their lower performance obtained in the WAKE condition. Older adults appear to rely more on initial higher

performance but become more vulnerable to incongruity between the cue triggering retrieval and the content of the action to perform. This may lead to potential difficulties to consolidate and recall unusual actions. Future studies may disentangle whether older adults do not anymore benefit from sleep in order to protect preexisting cognitive structures from excessive plasticity or whether this consolidation process is indeed impaired because of age-related changes in sleep architecture and microstructure, as it has already been described in the literature for classical episodic memory.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work was supported by grant from Observatoire B2V des mémoires for N.L.

References

Diekelmann S, Born J (2010) The memory function of sleep. *Nature Reviews Neuroscience* Available at: <https://doi.org/10.1038/nrn2762>.

Rauchs G, Desgranges B, Foret J, Eustache F (2005) The relationships between memory systems and sleep stages.. *J Sleep Res* 14:123–140.

Rauchs G, Feyers D, Landeau B, Bastin C, Luxen A, Maquet P, Collette F (2011) Sleep Contributes to the Strengthening of Some Memories Over Others Depending on Hippocampal Activity at Learning. *Journal of Neuroscience* 31:2563–2568 Available at: <https://doi.org/10.1523/jneurosci.3972-10.2011>.

van Dongen EV, Thielen J-W, Takashima A, Barth M, Fernández G (2012) Sleep Supports Selective Retention of Associative Memories Based on Relevance for Future Utilization Felmingham K, ed. *PLoS ONE* 7:e43426 Available at: <https://doi.org/10.1371/journal.pone.0043426>.

Wilhelm I, Diekelmann S, Molzow I, Ayoub A, Molle M, Born J (2011) Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *Journal of Neuroscience* 31:1563–1569 Available at: <https://doi.org/10.1523/jneurosci.3575-10.2011>.

Fischer S, Born J (2009) Anticipated reward enhances offline learning during sleep.. *Journal of Experimental Psychology: Learning Memory, and Cognition* 35:1586–1593 Available at: <https://doi.org/10.1037/a0017256>.

McDaniel GO M. A. & Einstein (2007) *Prospective Memory: An Overview and Synthesis of an Emerging Field*. SAGE Publications Inc. Available at: <https://doi.org/10.4135/2F9781452225913>.

Scullin MK, McDaniel MA (2010) Remembering to Execute a Goal. *Psychological Science*

21:1028–1035 Available at: <https://doi.org/10.1177%2F0956797610373373>.

Diekelmann S, Wilhelm I, Wagner U, Born J (2013) Sleep to Implement an Intention. *Sleep* 36:149–153 Available at: <https://doi.org/10.5665%2Fsleep.2322>.

Diekelmann S, Wilhelm I, Wagner U, Born J (2013) Sleep improves prospective remembering by facilitating spontaneous-associative retrieval processes.. *PLoS One* 8:e77621.

Einstein GO, McDaniel MA (1990) Normal aging and prospective memory.. *J Exp Psychol Learn Mem Cogn* 16:717–726.

Gonneaud J, Kalpouzos G, Bon L, Viader F, Eustache F, Desgranges B (2011) Distinct and shared cognitive functions mediate event- and time-based prospective memory impairment in normal ageing.. *Memory* 19:360–377.

McDaniel MA, Einstein GO (2011) The neuropsychology of prospective memory in normal aging: A componential approach. *Neuropsychologia* 49:2147–2155 Available at: <https://doi.org/10.1016%2Fj.neuropsychologia.2010.12.029>.

Uttl B (2008) Transparent Meta-Analysis of Prospective Memory and Aging Greene E, ed. *PLoS ONE* 3:e1568 Available at: <https://doi.org/10.1371%2Fjournal.pone.0001568>.

Lecouvey G, Gonneaud J, Piolino P, Madeleine S, Orriols E, Fleury P, Eustache F, Desgranges B (2017) Is binding decline the main source of the ageing effect on prospective memory? A ride in a virtual town.. *Socioaffect Neurosci Psychol* 7:1304610.

Aberle I, Rendell PG, Rose NS, McDaniel MA, Kliegel M (2010) The age prospective memory paradox: Young adults may not give their best outside of the lab.. *Developmental Psychology* 46:1444–1453 Available at: <https://doi.org/10.1037%2Fa0020718>.

Schnitzspahn KM, Ihle A, Henry JD, Rendell PG, Kliegel M (2010) The age-prospective memory-paradox: an exploration of possible mechanisms. *International Psychogeriatrics* 23:583–592 Avail-

able at: <https://doi.org/10.1017%2Fs1041610210001651>.

Mander BA, Winer JR, Walker MP (2017) Sleep and Human Aging. *Neuron* 94:19–36 Available at: <https://doi.org/10.1016%2Fj.neuron.2017.02.004>.

Backhaus J, Born J, Hoeckesfeld R, Fokuhl S, Hohagen F, Junghanns K (2007) Midlife decline in declarative memory consolidation is correlated with a decline in slow wave sleep.. *Learn Mem* 14:336–341.

Cherdiou M, Reynaud E, Uhrich J, Versace R, Mazza S (2014) Does age worsen sleep-dependent memory consolidation?. *J Sleep Res* 23:53–60.

Mander BA, Rao V, Lu B, Saletin JM, Lindquist JR, Ancoli-Israel S, Jagust W, Walker MP (2013) Prefrontal atrophy, disrupted NREM slow waves and impaired hippocampal-dependent memory in aging.. *Nat Neurosci* 16:357–364.

Aly M, Moscovitch M (2010) The effects of sleep on episodic memory in older and younger adults.. *Memory* 18:327–334.

Wilson JK, Baran B, Pace-Schott EF, Ivry RB, Spencer RM (2012) Sleep modulates word-pair learning but not motor sequence learning in healthy older adults.. *Neurobiol Aging* 33:991–1000.

Scullin M, Gao C, Krueger M, Ballard H, Bliwise D (2017) 0275 AGE EFFECTS ON REM SLEEP AND PROSPECTIVE MEMORY CONSOLIDATION. *Sleep* 40:A101–A101 Available at: <https://doi.org/10.1093%2Fsleepj%2Fzsx050.274>.

Buysse DJ, Reynolds CF3rd, Monk TH, Berman SR, Kupfer DJ (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.. *Psychiatry Res* 28:193–213.

Beck AT (1961) An Inventory for Measuring Depression. *Archives of General Psychiatry* 4:561 Available at: <https://doi.org/10.1001%2Farchpsyc.1961.01710120031004>.

Horne JA, Ostberg O (1976) A self-assessment questionnaire to determine morningness-

eveningness in human circadian rhythms.. *Int J Chronobiol* 4:97–110.

Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state. A practical method for grading the cognitive state of patients for the clinician.. *J Psychiatr Res* 12:189–198.

Akerstedt T, Gillberg M (1990) Subjective and objective sleepiness in the active individual.. *Int J Neurosci* 52:29–37.

Raven JC (1965) *The Mill Hill vocabulary scale*. Londre: H. K. Lewis.

Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA (1983) *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Ellis BW, Johns MW, Lancaster R, Raptopoulos P, Angelopoulos N, Priest RG (1981) The St. Mary's Hospital sleep questionnaire: a study of reliability.. *Sleep* 4:93–97.

Singmann H, Bolker B, Westfall J, Frederik A (2017) *afex: Analysis of Factorial Experiments*. R package version 0.17-8..

Lenth RV (2016) Least-Squares Means: TheRPackageIsmmeans. *Journal of Statistical Software* 69 Available at: <https://doi.org/10.18637%2Fjss.v069.i01>.

Silva GE, Goodwin JL, Sherrill DL, Arnold JL, Bootzin RR, Smith T, Walsleben JA, Baldwin CM, Quan SF (2007) Relationship between reported and measured sleep times: the sleep heart health study (SHHS).. *J Clin Sleep Med* 3:622–630.

Van DBJF, Van RFJ, Vos H, Tulen JH, Hofman A, Miedema HM, Neven AK, Tiemeier H (2008) Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons.. *J Sleep Res* 17:295–302.

Burgess PW, Quayle A, Frith CD (2001) Brain regions involved in prospective memory as determined by positron emission tomography.. *Neuropsychologia* 39:545–555.

Scullin MK, McDaniel MA, Shelton JT (2013) *The Dynamic Multiprocess Framework: evidence*

from prospective memory with contextual variability.. *Cogn Psychol* 67:55–71.

Gordon BA, Shelton JT, Bugg JM, McDaniel MA, Head D (2011) Structural correlates of prospective memory.. *Neuropsychologia* 49:3795–3800.

Gonneaud J, Piolino P, Lecouvey G, Madeleine S, Orriols E, Fleury P, Eustache F, Desgranges B (2012) Assessing prospective memory in young healthy adults using virtual reality. *Proc 9th Intl Conf Disability, Virtual Reality & Associated Technologies*:211–218.

Gilboa A, Marlatte H (2017) Neurobiology of Schemas and Schema-Mediated Memory.. *Trends Cogn Sci* 21:618–631.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG (2007) Schemas and memory consolidation.. *Science* 316:76–82.

van Kesteren MT, Ruitter DJ, Fernández G, Henson RN (2012) How schema and novelty augment memory formation.. *Trends Neurosci* 35:211–219.

van Kesteren MT, Rijpkema M, Ruitter DJ, Morris RG, Fernández G (2014) Building on prior knowledge: schema-dependent encoding processes relate to academic performance.. *J Cogn Neurosci* 26:2250–2261.

van Kesteren MT, Beul SF, Takashima A, Henson RN, Ruitter DJ, Fernández G (2013) Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent.. *Neuropsychologia* 51:2352–2359.

Landmann N, Kuhn M, Piosczyk H, Feige B, Baglioni C, Spiegelhalder K, Frase L, Riemann D, Sterr A, Nissen C (2014) The reorganisation of memory during sleep.. *Sleep Med Rev* 18:531–541.

Lewis PA, Durrant SJ (2011) Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences* 15:343–351 Available at: <https://doi.org/10.1016%2Fj.tics.2011.06.004>.

Hennies N, Ralph MAL, Kempkes M, Cousins JN, Lewis PA (2016) Sleep Spindle Density Predicts the Effect of Prior Knowledge on Memory Consolidation. *The Journal of Neuroscience* 36:3799–3810 Available at: <https://doi.org/10.1523%2Fjneurosci.3162-15.2016>.

Durrant SJ, Cairney SA, McDermott C, Lewis PA (2015) Schema-conformant memories are preferentially consolidated during REM sleep.. *Neurobiol Learn Mem* 122:41–50.

Groch S, Schreiner T, Rasch B, Huber R, Wilhelm I (2017) Prior knowledge is essential for the beneficial effect of targeted memory reactivation during sleep. *Scientific Reports* 7:39763 Available at: <https://doi.org/10.1038%2Fsrep39763>.

Mander BA, Rao V, Lu B, Saletin JM, Lindquist JR, Ancoli-Israel S, Jagust W, Walker MP (2013) Prefrontal atrophy disrupted NREM slow waves and impaired hippocampal-dependent memory in aging. *Nature Neuroscience* 16:357–364 Available at: <https://doi.org/10.1038%2Fnn.3324>.

Helfrich RF, Mander BA, Jagust WJ, Knight RT, Walker MP (2018) Old Brains Come Uncoupled in Sleep: Slow Wave-Spindle Synchrony Brain Atrophy, and Forgetting. *Neuron* 97:221–230.e4 Available at: <https://doi.org/10.1016%2Fj.neuron.2017.11.020>.

Stern Y (2012) Cognitive reserve in ageing and Alzheimer's disease.. *Lancet Neurol* 11:1006–1012.

(2014). In:

- Akselrod, S, D Gordon, F. Ubel, et al. (1981). „Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control“. In: *Science* 213.4504, pp. 220–222.
- Alberini, Cristina M. (2005). „Mechanisms of memory stabilization: are consolidation and reconsolidation similar or distinct processes?“. In: *Trends in Neurosciences* 28.1, pp. 51–56.
- Aldao, Amelia, Susan Nolen-Hoeksema, and Susanne Schweizer (2010). „Emotion-regulation strategies across psychopathology: A meta-analytic review“. In: *Clinical Psychology Review* 30.2, pp. 217–237.
- Anderson, Kristopher L., Rajasimhan Rajagovindan, Georges A. Ghacibeh, Kimford J. Meador, and Mingzhou Ding (2009). „Theta Oscillations Mediate Interaction between Prefrontal Cortex and Medial Temporal Lobe in Human Memory“. In: *Cerebral Cortex* 20.7, pp. 1604–1612.
- Anderson, Michael C. (2003). „Rethinking interference theory: Executive control and the mechanisms of forgetting“. In: *Journal of Memory and Language* 49.4, pp. 415–445.
- (2004). „Neural Systems Underlying the Suppression of Unwanted Memories“. In: *Science* 303.5655, pp. 232–235.
- Anderson, Michael C. and Collin Green (2001). „Suppressing unwanted memories by executive control“. In: *Nature* 410.6826, pp. 366–369.
- Anderson, Michael C., RA Bjork, and EL Bjork (1994). „Remembering can cause forgetting: retrieval dynamics in long-term memory.“ In: *J Exp Psychol Learn Mem Cogn* 20, pp. 1063–87.
- Anderson, Michael C., Julia Reinholz, Brice A. Kuhl, and Ulrich Mayr (2011). „Intentional suppression of unwanted memories grows more difficult as we age.“ In: *Psychology and Aging* 26.2, pp. 397–405.
- Anderson, Michael C., Jamie G. Bunce, and Helen Barbas (2016). „Prefrontal–hippocampal pathways underlying inhibitory control over memory“. In: *Neurobiology of Learning and Memory* 134, pp. 145–161.
- Assmann, Jan (2011). „Communicative and Cultural Memory“. In: *Cultural Memories*. Springer Netherlands, pp. 15–27.
- Assmann, Jan and John Czaplicka (1995). „Collective Memory and Cultural Identity“. In: *New German Critique* 65, p. 125.
- Babapoor-Farrokhran, Sahand, Martin Vinck, Thilo Womelsdorf, and Stefan Everling (2017). „Theta and beta synchrony coordinate frontal eye fields and anterior cingulate cortex during sensorimotor mapping“. In: *Nature Communications*, p. 13967.
- Babo-Rebelo, Mariana, Nicolai Wolpert, Claude Adam, Dominique Hasboun, and Catherine Tallon-Baudry (2016a). „Is the cardiac monitoring function related to the self in both the default network and right anterior insula?“. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1708, p. 20160004.
- Babo-Rebelo, Mariana, Craig G. Richter, and Catherine Tallon-Baudry (2016b). „Neural Responses to Heartbeats in the Default Network Encode the Self in Spontaneous Thoughts“. In: *The Journal of Neuroscience* 36.30, pp. 7829–7840.

- Backus, Alexander R., Jan-Mathijs Schoffelen, Szabolcs Szebényi, Simon Hanslmayr, and Christian F. Doeller (2016). „Hippocampal-Prefrontal Theta Oscillations Support Memory Integration“. In: *Current Biology* 26.4, pp. 450–457.
- Barnier, Amanda J. and John Sutton (2008). „From individual to collective memory: Theoretical and empirical perspectives“. In: *Memory* 16.3, pp. 177–182.
- Barron, H.C., T.P. Vogels, U.E. Emir, et al. (2016). „Unmasking Latent Inhibitory Connections in Human Cortex to Reveal Dormant Cortical Memories“. In: *Neuron* 90.1, pp. 191–203.
- Barron, Helen C., Tim P. Vogels, Timothy E. Behrens, and Mani Ramaswami (2017). „Inhibitory engrams in perception and memory“. In: *Proceedings of the National Academy of Sciences*, p. 201701812.
- Bartlett, Frederic (1932/1995). *Remembering : a study in experimental and social psychology*. Cambridge New York: Cambridge University Press.
- Bavel, Jay J. Van and Andrea Pereira (2018). „The Partisan Brain: An Identity-Based Model of Political Belief“. In: *Trends in Cognitive Sciences* 22.3, pp. 213–224.
- Beauchaine, Theodore P., Lisa Gatzke-Kopp, and Hilary K. Mead (2007). „Polyvagal Theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence“. In: *Biological Psychology* 74.2, pp. 174–184.
- Benoit, Roland G. and Michael C. Anderson (2012). „Opposing Mechanisms Support the Voluntary Forgetting of Unwanted Memories“. In: *Neuron* 76.2, pp. 450–460.
- 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“. In: *Journal of Cognitive Neuroscience* 27.1, pp. 96–111.
- Bergson, Henri (2012). *Matière et mémoire : essai sur la relation du corps à l'esprit*. Paris: Flammarion.
- Bernard, Claude (1865). „étude sur la physiologie du cœur“. In: *Revue des Deux Mondes* T.56.
- Bietti, Lucas M., Otilie Tilston, and Adrian Bangerter (2018). „Storytelling as Adaptive Collective Sensemaking“. In: *Topics in Cognitive Science*.
- Billman, George E. (2011). „Heart Rate Variability ? A Historical Perspective“. In: *Frontiers in Physiology* 2.
- Borges, Jorge (1993). *Fictions*. Paris: Gallimard.
- Bray, Patrick M. (2013). „Forgetting the Madeleine“. In: *Progress in Brain Research*. Elsevier, pp. 41–53.
- Brewin, Chris R. (2014). „Episodic memory perceptual memory, and their interaction: Foundations for a theory of posttraumatic stress disorder.“ In: *Psychological Bulletin* 140.1, pp. 69–97.
- Brewin, Chris R., James D. Gregory, Michelle Lipton, and Neil Burgess (2010). „Intrusive images in psychological disorders: Characteristics neural mechanisms, and treatment implications.“ In: *Psychological Review* 117.1, pp. 210–232.
- Brown, Adam D., Nicole Kouri, and William Hirst (2012). „Memory's Malleability: Its Role in Shaping Collective Memory and Social Identity“. In: *Frontiers in Psychology* 3.
- Bruyère, Jean de la (1688/1995). *Les caractères*. Paris: Librairie générale française.
- Burke, J. F., A. D. Sharan, M. R. Sperling, et al. (2014). „Theta and High-Frequency Activity Mark Spontaneous Recall of Episodic Memories“. In: *Journal of Neuroscience* 34.34, pp. 11355–11365.
- Castel, P-H. (2006). *A quoi résiste la psychanalyse*. Paris: Presses universitaires de France.

- Catarino, Ana, Charlotte S. Küpper, Aliza Werner-Seidler, Tim Dalgleish, and Michael C. Anderson (2015). „Failing to Forget“. In: *Psychological Science* 26.5, pp. 604–616.
- Cavanagh, James F. and Michael J. Frank (2014). „Frontal theta as a mechanism for cognitive control“. In: *Trends in Cognitive Sciences* 18.8, pp. 414–421.
- Chalmers, John A., Daniel S. Quintana, Maree J.-Anne Abbott, and Andrew H. Kemp (2014). „Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis“. In: *Frontiers in Psychiatry* 5.
- Chen, Janice, Yuan Chang Leong, Christopher J Honey, et al. (2016). „Shared memories reveal shared structure in neural activity across individuals“. In: *Nature Neuroscience* 20.1, pp. 115–125.
- Chmielewski, Witold X., Moritz Mückschel, Gabriel Dippel, and Christian Beste (2015). „Concurrent information affects response inhibition processes via the modulation of theta oscillations in cognitive control networks“. In: *Brain Structure and Function* 221.8, pp. 3949–3961.
- Clarini, J. (2015). „« Traité sur la tolérance » et « Paris est une fête », best-sellers inattendus“. In: *Le monde*.
- Clark, A. and D. Chalmers (1998). „The Extended Mind“. In: *Analysis* 58.1, pp. 7–19.
- Clouter, Andrew, Kimron L. Shapiro, and Simon Hanslmayr (2017). „Theta Phase Synchronization Is the Glue that Binds Human Associative Memory“. In: *Current Biology* 27.20, 3143–3148.e6.
- Cohen, Michael X (2017). „Where Does EEG Come From and What Does It Mean?“ In: *Trends in Neurosciences* 40.4, pp. 208–218.
- Coman, Alin, David Manier, and William Hirst (2009). „Forgetting the Unforgettable Through Conversation“. In: *Psychological Science* 20.5, pp. 627–633.
- Coman, Alin, Charles B. Stone, Emanuele Castano, and William Hirst (2014). „Justifying Atrocities“. In: *Psychological Science* 25.6, pp. 1281–1285.
- Coman, Alin, Ida Momennejad, Rae D. Drach, and Andra Geana (2016). „Mnemonic convergence in social networks: The emergent properties of cognition at a collective level“. In: *Proceedings of the National Academy of Sciences* 113.29, pp. 8171–8176.
- Crews, Frederick (1995). *The memory wars : Freud's legacy in dispute*. New York: New York Review of Books.
- Critchley, Hugo D and Sarah N Garfinkel (2018). „The influence of physiological signals on cognition“. In: *Current Opinion in Behavioral Sciences* 19, pp. 13–18.
- Cuc, Alexandru, Jonathan Koppel, and William Hirst (2007). „Silence Is Not Golden“. In: *Psychological Science* 18.8, pp. 727–733.
- Damasio, Antonio (1999). *The feeling of what happens : body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Darwin, Charles (1872). *The expression of the emotions in man and animals*. John Murray.
- Daselaar, S. M., H. J. Rice, D. L. Greenberg, et al. (2007). „The Spatiotemporal Dynamics of Autobiographical Memory: Neural Correlates of Recall Emotional Intensity, and Reliving“. In: *Cerebral Cortex* 18.1, pp. 217–229.
- Davis, Ronald L. and Yi Zhong (2017). „The Biology of Forgetting—A Perspective“. In: *Neuron* 95.3, pp. 490–503.
- De, Jong MJ and DC Randall (2005). „Heart rate variability analysis in the assessment of autonomic function in heart failure.“ In: *J Cardiovasc Nurs* 20, 186–95; quiz 196–7.

- „Decoding the Dynamics of Conscious Perception: The Temporal Generalization Method“ (2016). In:
- Depue, B. E., T. Curran, and M. T. Banich (2007). „Prefrontal Regions Orchestrate Suppression of Emotional Memories via a Two-Phase Process“. In: *Science* 317.5835, pp. 215–219.
- Depue, B. E., J. M. Orr, H. R. Smolker, F. Naaz, and M. T. Banich (2015). „The Organization of Right Prefrontal Networks Reveals Common Mechanisms of Inhibitory Regulation Across Cognitive Emotional, and Motor Processes“. In: *Cerebral Cortex* 26.4, pp. 1634–1646.
- Depue, Brendan E., Marie T. Banich, and Tim Curran (2006). „Suppression of Emotional and Nonemotional Content in Memory“. In: *Psychological Science* 17.5, pp. 441–447.
- Dessalles (2007). *Why we talk : the evolutionary origins of language*. Oxford New York: Oxford University Press.
- Diekelmann, Susanne and Jan Born (2010). „The memory function of sleep“. In: *Nature Reviews Neuroscience* 11.2, pp. 114–126.
- Dieler, Alica C., Michael M. Plichta, T. Dresler, and Andreas J. Fallgatter (2010). „Suppression of emotional words in the Think/No-Think paradigm investigated with functional near-infrared spectroscopy“. In: *International Journal of Psychophysiology* 78.2, pp. 129–135.
- Disner, Seth G., Christopher G. Beevers, Emily A. P. Haigh, and Aaron T. Beck (2011). „Neural mechanisms of the cognitive model of depression“. In: *Nature Reviews Neuroscience* 12.8, pp. 467–477.
- Doss, Manoj K., Jamila K. Picart, and David A. Gallo (2018). „The Dark Side of Context: Context Reinstatement Can Distort Memory“. In: *Psychological Science*, p. 095679761774953.
- Dosse, François (2010). *Renaissance de l'événement : un défi pour l'historien : entre sphinx et phénix*. Paris: Presses universitaires de France.
- Dwyer, Dominic B., Peter Falkai, and Nikolaos Koutsouleris (2018). „Machine Learning Approaches for Clinical Psychology and Psychiatry“. In: *Annual Review of Clinical Psychology* 14.1, pp. 91–118.
- Ebbinghaus, Hermann (1885/2010). *La mémoire : recherches de psychologie expérimentale*. Paris: L'Harmattan.
- (2011). *Über das Gedächtnis : Untersuchungen zur experimentellen Psychologie*. Darmstadt: WBG, Wiss. Buchges.
- Eco, Umberto and Marilyn Migiel (1988). „An Ars Oblivionalis? Forget It!“ In: *PMLA* 103.3, p. 254.
- Edelson, M., T. Sharot, R. J. Dolan, and Y. Dudai (2011). „Following the Crowd: Brain Substrates of Long-Term Memory Conformity“. In: *Science* 333.6038, pp. 108–111.
- Engen, Haakon G. and Michael C. Anderson (2018). „Memory Control: A Fundamental Mechanism of Emotion Regulation“. In: *Trends in Cognitive Sciences*.
- Eschmann, Kathrin C.J., Regine Bader, and Axel Mecklinger (2018). „Topographical differences of frontal-midline theta activity reflect functional differences in cognitive control abilities“. In: *Brain and Cognition* 123, pp. 57–64.
- Etkin, Amit, Christian Büchel, and James J. Gross (2015). „The neural bases of emotion regulation“. In: *Nature Reviews Neuroscience* 16.11, pp. 693–700.
- Eustache, Francis (2014). *Mémoire et oubli*. Paris: Éd. le Pommier.
- Fell, Juergen and Nikolai Axmacher (2011). „The role of phase synchronization in memory processes“. In: *Nature Reviews Neuroscience* 12.2, pp. 105–118.

- Folstein, Jonathan R. and Cyma Van Petten (2007). „Influence of cognitive control and mismatch on the N2 component of the ERP: A review“. In: *Psychophysiology* 0.0, 070915195953001–???
- Frank, M. J., C. Gagne, E. Nyhus, et al. (2015). „fMRI and EEG Predictors of Dynamic Decision Parameters during Human Reinforcement Learning“. In: *Journal of Neuroscience* 35.2, pp. 485–494.
- Frankland, Paul W. and Bruno Bontempi (2005). „The organization of recent and remote memories“. In: *Nature Reviews Neuroscience* 6.2, pp. 119–130.
- Friston, Karl (2010). „The free-energy principle: a unified brain theory?“. In: *Nature Reviews Neuroscience* 11.2, pp. 127–138.
- Fuentemilla, L., G.R. Barnes, E. Düzel, and B. Levine (2014). „Theta oscillations orchestrate medial temporal lobe and neocortex in remembering autobiographical memories“. In: *NeuroImage* 85, pp. 730–737.
- Gagnepain, Pierre, Richard N. Henson, and Michael C. Anderson (2014). „Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition“. In: *Proceedings of the National Academy of Sciences* 111.13, E1310–E1319.
- Gagnepain, Pierre, Justin Hulbert, and Michael C. Anderson (2017). „Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories“. In: *The Journal of Neuroscience* 37.27, pp. 6423–6441.
- Gagnepain, Pierre, Thomas Vallée, Serge Heiden, et al. (forthcoming). *Collective memory shapes the organization of individual memories in the medial prefrontal cortex*.
- Garfinkel, Sarah N. and Hugo D. Critchley (2016). „Threat and the Body: How the Heart Supports Fear Processing“. In: *Trends in Cognitive Sciences* 20.1, pp. 34–46.
- Gazzellini, Simone, Maria Dettori, Francesca Amadori, et al. (2016). „Association between Attention and Heart Rate Fluctuations in Pathological Worriers“. In: *Frontiers in Human Neuroscience* 10.
- Ghosh, Vanessa E. and Asaf Gilboa (2014). „What is a memory schema? A historical perspective on current neuroscience literature“. In: *Neuropsychologia* 53, pp. 104–114.
- Gilboa, Asaf and Hannah Marlatte (2017). „Neurobiology of Schemas and Schema-Mediated Memory“. In: *Trends in Cognitive Sciences* 21.8, pp. 618–631.
- Gillie, Brandon L. and Julian F. Thayer (2014). „Individual differences in resting heart rate variability and cognitive control in posttraumatic stress disorder“. In: *Frontiers in Psychology* 5.
- Gillie, Brandon L., Michael W. Vasey, and Julian F. Thayer (2013). „Heart Rate Variability Predicts Control Over Memory Retrieval“. In: *Psychological Science* 25.2, pp. 458–465.
- (2015). „Individual differences in resting heart rate variability moderate thought suppression success“. In: *Psychophysiology* 52.9, pp. 1149–1160.
- Goldin, Philippe R., Kateri McRae, Wiveka Ramel, and James J. Gross (2008). „The Neural Bases of Emotion Regulation: Reappraisal and Suppression of Negative Emotion“. In: *Biological Psychiatry* 63.6, pp. 577–586.
- Grootswagers, Tijn, Susan G. Wardle, and Thomas A. Carlson (2017). „Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data“. In: *Journal of Cognitive Neuroscience* 29.4, pp. 677–697.
- Gross, James J. (2015). „The Extended Process Model of Emotion Regulation: Elaborations Applications, and Future Directions“. In: *Psychological Inquiry* 26.1, pp. 130–137.

- Gross, James J. and Hooria Jazaieri (2014). „Emotion Emotion Regulation, and Psychopathology“. In: *Clinical Psychological Science* 2.4, pp. 387–401.
- Guskjolen, Axel, Justin W. Kenney, Juan de la Parra, et al. (2018). „Recovery of Lost Infant Memories in Mice“. In: *Current Biology* 28.14, 2283–2290.e3.
- Halbwachs, Maurice (1925). *Les cadres sociaux de la mémoire*. Paris: Albin Michel.
- Hanslmayr, Simon, Bernhard Pastötter, Karl-Heinz Bäuml, et al. (2008). „The Electrophysiological Dynamics of Interference during the Stroop Task“. In: *Journal of Cognitive Neuroscience* 20.2, pp. 215–225.
- Hanslmayr, Simon, Tobias Staudigl, Alp Aslan, and Karl-Heinz Bäuml (2010). „Theta oscillations predict the detrimental effects of memory retrieval“. In: *Cognitive Affective, & Behavioral Neuroscience* 10.3, pp. 329–338.
- Hanslmayr, Simon, Tobias Staudigl, and Marie-Christin Fellner (2012). „Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis“. In: *Frontiers in Human Neuroscience* 6.
- Hasselmo, Michael E., Clara Bodelón, and Bradley P. Wyble (2002). „A Proposed Function for Hippocampal Theta Rhythm: Separate Phases of Encoding and Retrieval Enhance Reversal of Prior Learning“. In: *Neural Computation* 14.4, pp. 793–817.
- Haynes, John-Dylan (2015). „A Primer on Pattern-Based Approaches to fMRI: Principles Pitfalls, and Perspectives“. In: *Neuron* 87.2, pp. 257–270.
- Head, Henry and Gordon Holmes (1911). „Sensory disturbances from cerebral lesions“. In: *Brain* 34.2-3, pp. 102–254.
- Hebart, Martin N. and Chris I. Baker (2018). „Deconstructing multivariate decoding for the study of brain function“. In: *NeuroImage* 180, pp. 4–18.
- Helfrich, Randolph F. and Robert T. Knight (2016). „Oscillatory Dynamics of Prefrontal Cognitive Control“. In: *Trends in Cognitive Sciences* 20.12, pp. 916–930.
- Hemingway, Ernest (1964/2012). *Paris est une fête*. Paris: Gallimard.
- Henson, Richard N. and Pierre Gagnepain (2010). „Predictive interactive multiple memory systems“. In: *Hippocampus* 20.11, pp. 1315–1326.
- Herbert, Cornelia and Stefan Sütterlin (2012). „Do Not Respond! Doing the Think/No-Think and Go/No-Go Tasks Concurrently Leads to Memory Impairment of Unpleasant Items during Later Recall“. In: *Frontiers in Psychology* 3.
- Hertel, Paula and Leda McDaniel (2010). „The suppressive power of positive thinking: Aiding suppression-induced forgetting in repressive coping“. In: *Cognition & Emotion* 24.7, pp. 1239–1249.
- Hertel, Paula T. and Melissa Gerstle (2003). „Depressive Deficits in Forgetting“. In: *Psychological Science* 14.6, pp. 573–578.
- Herz, Rachel S. and Jonathan W. Schooler (2002). „A Naturalistic Study of Autobiographical Memories Evoked by Olfactory and Visual Cues: Testing the Proustian Hypothesis“. In: *The American Journal of Psychology* 115.1, p. 21.
- Heusser, Andrew C, David Poeppel, Youssef Ezzyat, and Lila Davachi (2016). „Episodic sequence memory is supported by a theta–gamma phase code“. In: *Nature Neuroscience* 19.10, pp. 1374–1380.
- Hirst, William and David Manier (2008). „Towards a psychology of collective memory“. In: *Memory* 16.3, pp. 183–200.
- Hirst, William, Jeremy K. Yamashiro, and Alin Coman (2018). „Collective Memory from a Psychological Perspective“. In: *Trends in Cognitive Sciences* 22.5, pp. 438–451.

- Hon, EH and ST Lee (1963). „ELECTRONIC EVALUATION OF THE FETAL HEART RATE. VIII. PATTERNS PRECEDING FETAL DEATH, FURTHER OBSERVATIONS.“ In: *Am J Obstet Gynecol* 87, pp. 814–26.
- Horikawa, T., M. Tamaki, Y. Miyawaki, and Y. Kamitani (2013). „Neural Decoding of Visual Imagery During Sleep“. In: *Science* 340.6132, pp. 639–642.
- Hu, Xiaoqing, Zara M. Bergström, Pierre Gagnepain, and Michael C. Anderson (2017). „Suppressing Unwanted Memories Reduces Their Unintended Influences“. In: *Current Directions in Psychological Science* 26.2, pp. 197–206.
- Hulbert, Justin C. and Michael C. Anderson (2018). „What doesn't kill you makes you stronger: Psychological trauma and its relationship to enhanced memory control.“ In: *Journal of Experimental Psychology: General*.
- Hulbert, Justin C., Richard N. Henson, and Michael C. Anderson (2016). „Inducing amnesia through systemic suppression“. In: *Nature Communications* 7, p. 11003.
- Höping, Winfried and Renate de Jong-Meyer (2003). „Differentiating unwanted intrusive thoughts from thought suppression: what does the White Bear Suppression Inventory measure?“ In: *Personality and Individual Differences* 34.6, pp. 1049–1055.
- James, William (1884). „II.—WHAT IS AN EMOTION ?“ In: *Mind* os-IX.34, pp. 188–205.
- (1892). *Psychology : the briefer course*. Mineola, N.Y: Dover Publications.
- Janet, P. (1926). *De l'angoisse à l'extase*. Paris, Société Pierre Janet et Laboratoire de psychologie pathologique de la Sorbonne.
- Jirsa, Viktor and Viktor Müller (2013). „Cross-frequency coupling in real and virtual brain networks“. In: *Frontiers in Computational Neuroscience* 7.
- John, Oliver P. and James J. Gross (2004). „Healthy and Unhealthy Emotion Regulation: Personality Processes Individual Differences, and Life Span Development“. In: *Journal of Personality* 72.6, pp. 1301–1334.
- Joormann, Jutta, Paula T. Hertel, Faith Brozovich, and Ian H. Gotlib (2005). „Remembering the Good Forgetting the Bad: Intentional Forgetting of Emotional Material in Depression.“ In: *Journal of Abnormal Psychology* 114.4, pp. 640–648.
- Joormann, Jutta, Paula T. Hertel, Joelle LeMoult, and Ian H. Gotlib (2009). „Training forgetting of negative material in depression.“ In: *Journal of Abnormal Psychology* 118.1, pp. 34–43.
- Kamitani, Yukiyasu and Frank Tong (2005). „Decoding the visual and subjective contents of the human brain“. In: *Nature Neuroscience* 8.5, pp. 679–685.
- Karakaya, Osman, Irfan Barutcu, Dayimi Kaya, et al. (2007). „Acute Effect of Cigarette Smoking on Heart Rate Variability“. In: *Angiology* 58.5, pp. 620–624.
- Kerrén, Casper, Juan Linde-Domingo, Simon Hanslmayr, and Maria Wimber (2018). „An Optimal Oscillatory Phase for Pattern Reactivation during Memory Retrieval“. In: *Current Biology*.
- Kesteren, Marlieke T.R. van, Dirk J. Ruiters, Guillén Fernández, and Richard N. Henson (2012). „How schema and novelty augment memory formation“. In: *Trends in Neurosciences* 35.4, pp. 211–219.
- Kesteren, Marlieke T.R. van, Sarah F. Beul, Atsuko Takashima, et al. (2013). „Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent“. In: *Neuropsychologia* 51.12, pp. 2352–2359.
- Ketz, Nicholas, Randal C. O'Reilly, and Tim Curran (2014). „Classification aided analysis of oscillatory signatures in controlled retrieval“. In: *NeuroImage* 85, pp. 749–760.

- King, J-R. and S. Dehaene (2014). „Characterizing the dynamics of mental representations: the temporal generalization method“. In: *Trends in Cognitive Sciences* 18.4, pp. 203–210.
- King, Jean-Rémi, Niccolo Pescetelli, and Stanislas Dehaene (2016). „Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information“. In: *Neuron* 92.5, pp. 1122–1134.
- Knill, David C. and Alexandre Pouget (2004). „The Bayesian brain: the role of uncertainty in neural coding and computation“. In: *Trends in Neurosciences* 27.12, pp. 712–719.
- Kreibig, Sylvia D. (2010). „Autonomic nervous system activity in emotion: A review“. In: *Biological Psychology* 84.3, pp. 394–421.
- Kriegeskorte, Nikolaus (2008). „Representational similarity analysis – connecting the branches of systems neuroscience“. In: *Frontiers in Systems Neuroscience*.
- Kriegeskorte, Nikolaus and Peter Bandettini (2007). „Analyzing for information not activation, to exploit high-resolution fMRI“. In: *NeuroImage* 38.4, pp. 649–662.
- Lambert, Anthony J., Kimberly S. Good, and Ian J. Kirk (2010). „Testing the repression hypothesis: Effects of emotional valence on memory suppression in the think – No think task“. In: *Consciousness and Cognition* 19.1, pp. 281–293.
- Langley, JN (1921). *The Autonomic Nervous System*. Heffer & Sons; Cambridge, England.
- Lavabre, Marie-Claire (2013). „A propos de l’oubli dans les réflexions sur la mémoire collective ou sociale“. In: *Mémoire et mémorialisation*. Ed. by Ed Berenson et Denis Peschanski. De l’absence à la représentation. Hermann.
- Lee, Francis LF, Joseph Man Chan, and Dennis KK Leung (2017). „When a historical analogy fails: Current political events and collective memory contestation in the news“. In: *Memory Studies*, p. 175069801770380.
- Lee, Sue-Hyun, Dwight J. Kravitz, and Chris I. Baker (2012). „Disentangling visual imagery and perception of real-world objects“. In: *NeuroImage* 59.4, pp. 4064–4073.
- Lega, Bradley, John Burke, Joshua Jacobs, and Michael J. Kahana (2014). „Slow-Theta-to-Gamma Phase–Amplitude Coupling in Human Hippocampus Supports the Formation of New Episodic Memories“. In: *Cerebral Cortex* 26.1, pp. 268–278.
- Legrand, Nicolas, Pierre Gagnepain, Denis Peschanski, and Francis Eustache (2015). „Neurosciences et mémoires collectives : les schémas entre cerveau sociétés et cultures“. In: *Biologie Aujourd’hui* 209.3, pp. 273–286.
- Lehrer, Jonah (2007). *Proust was a neuroscientist*. Boston: Houghton Mifflin Co.
- Levy, Benjamin J. and Michael C. Anderson (2008). „Individual differences in the suppression of unwanted memories: The executive deficit hypothesis“. In: *Acta Psychologica* 127.3, pp. 623–635.
- Lischke, Alexander, Daniela Lemke, Jörg Neubert, Alfons O. Hamm, and Martin Lotze (2017). „Inter-individual differences in heart rate variability are associated with inter-individual differences in mind-reading“. In: *Scientific Reports* 7.1.
- Liu, Yunzhe, Wanjun Lin, Chao Liu, et al. (2016). „Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories“. In: *Nature Communications* 7, p. 13375.
- Luft, Caroline Di Bernardi and Joydeep Bhattacharya (2015). „Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates“. In: *Scientific Reports* 5.1.
- Lundqvist, Mikael, Jonas Rose, Pawel Herman, et al. (2016). „Gamma and Beta Bursts Underlie Working Memory“. In: *Neuron* 90.1, pp. 152–164.

- Lundqvist, Mikael, Pawel Herman, Melissa R. Warden, Scott L. Brincat, and Earl K. Miller (2018). „Gamma and beta bursts during working memory readout suggest roles in its volitional control“. In: *Nature Communications* 9.1.
- Lyoo, In Kyoon (2011). „The Neurobiological Role of the Dorsolateral Prefrontal Cortex in Recovery From Trauma“. In: *Archives of General Psychiatry* 68.7, p. 701.
- MacLeod, Colin M. (1975). „Long-term recognition and recall following directed forgetting.“ In: *Journal of Experimental Psychology: Human Learning and Memory* 1.3, pp. 271–279.
- Magee, Joshua C., K. Paige Harden, and Bethany A. Teachman (2012). „Psychopathology and thought suppression: A quantitative review“. In: *Clinical Psychology Review* 32.3, pp. 189–201.
- Mahr, Johannes B. and Gergely Csibra (2017). „Why do we remember? The communicative function of episodic memory“. In: *Behavioral and Brain Sciences* 41.
- Maswood, Raeya, Anne S. Rasmussen, and Suparna Rajaram (2018). „Collaborative remembering of emotional autobiographical memories: Implications for emotion regulation and collective memory.“ In: *Journal of Experimental Psychology: General*.
- McCorry, Laurie Kelly (2007). „Physiology of the Autonomic Nervous System“. In: *American Journal of Pharmaceutical Education* 71.4, p. 78.
- Merck, Clinton, Meymune N Topcu, and William Hirst (2016). „Collective mental time travel: Creating a shared future through our shared past“. In: *Memory Studies* 9.3. Ed. by Daniel L Schacter and Michael Welker, pp. 284–294.
- Meshi, Dar, Diana I. Tamir, and Hauke R. Heekeren (2015). „The Emerging Neuroscience of Social Media“. In: *Trends in Cognitive Sciences* 19.12, pp. 771–782.
- Mihailova, Stella and Laura Jobson (2018). „Association between intrusive negative autobiographical memories and depression: A meta-analytic investigation“. In: *Clinical Psychology & Psychotherapy* 25.4, pp. 509–524.
- Minassian, Arpi, Adam X. Maihofer, Dewleen G. Baker, et al. (2015). „Association of Pre-deployment Heart Rate Variability With Risk of Postdeployment Posttraumatic Stress Disorder in Active-Duty Marines“. In: *JAMA Psychiatry* 72.10, p. 979.
- Morawetz, Carmen, Stefan Bode, Birgit Derntl, and Hauke R. Heekeren (2017). „The effect of strategies goals and stimulus material on the neural mechanisms of emotion regulation: A meta-analysis of fMRI studies“. In: *Neuroscience & Biobehavioral Reviews* 72, pp. 111–128.
- Murray, Brendan D., Keely A. Muscatell, and Elizabeth A. Kensinger (2011). „Effects of emotion and age on performance during a think/no-think memory task.“ In: *Psychology and Aging* 26.4, pp. 940–955.
- Müller, Klaus-Robert, Michael Tangermann, Guido Dornhege, et al. (2008). „Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring“. In: *Journal of Neuroscience Methods* 167.1, pp. 82–90.
- Naccache, Lionel (2006). *Le nouvel inconscient : Freud, Christophe Colomb des neurosciences*. Paris: O. Jacob.
- Nietzsche, Friedrich (1874). *Seconde considération intempestive : de l'utilité et de l'inconvénient des études historiques pour la vie*. Paris: Flammarion.
- (1881/2012). *Aurore : pensées sur les préjugés moraux*. Paris: Flammarion.
- Nigbur, Roland, Galina Ivanova, and Birgit Stürmer (2011). „Theta power as a marker for cognitive interference“. In: *Clinical Neurophysiology* 122.11, pp. 2185–2194.

- Norman, Kenneth A., Sean M. Polyn, Greg J. Detre, and James V. Haxby (2006). „Beyond mind-reading: multi-voxel pattern analysis of fMRI data“. In: *Trends in Cognitive Sciences* 10.9, pp. 424–430.
- Nunan, David, Gavin R. H. Sandercock, and David A. Brodie (2010). „A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults“. In: *Pacing and Clinical Electrophysiology* 33.11, pp. 1407–1417.
- Nyhus, Erika and Tim Curran (2010). „Functional role of gamma and theta oscillations in episodic memory“. In: *Neuroscience & Biobehavioral Reviews* 34.7, pp. 1023–1035.
- Nørby, S, M Lange, and A Larsen (2010). „Forgetting to forget: on the duration of voluntary suppression of neutral and emotional memories.“ In: *Acta Psychol (Amst)* 133, pp. 73–80.
- Pagani, M, F Lombardi, S Guzzetti, et al. (1984). „Power spectral density of heart rate variability as an index of sympatho-vagal interaction in normal and hypertensive subjects.“ In: *J Hypertens Suppl* 2, S383–5.
- Panksepp, Jaak (1998). *Affective neuroscience : the foundations of human and animal emotions*. New York: Oxford University Press.
- Panksepp, Jaak and Mark Solms (2012). „What is neuropsychanalysis? Clinically relevant studies of the minded brain“. In: *Trends in Cognitive Sciences* 16.1, pp. 6–8.
- Park, Joo Eon, Ji Yeon Lee, Suk-Hoon Kang, et al. (2017). „Heart rate variability of chronic posttraumatic stress disorder in the Korean veterans“. In: *Psychiatry Research* 255, pp. 72–77.
- Pedregosa, F., G. Varoquaux, A. Gramfort, et al. (2011). „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.
- Pennebaker, James W. and Amy L. Gonzales (2009). „Making History: Social and Psychological Processes Underlying Collective Memory“. In: *Memory in Mind and Culture*. Ed. by Pascal Boyer and James V. Wertsch. Cambridge University Press, pp. 171–193.
- Peschanski, D. and B. Sion (2018). *La Vérité du témoin: Mémoire et mémorialisation vol. 2*. Hermann.
- Peschanski, Denis (2013). *Mémoire et mémorialisation*. Paris: Hermann.
- Platon (-315/2016). *La République*. Paris: GF Flammarion.
- Porges, Stephen W. (1995). „Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A Polyvagal Theory.“ In: *Psychophysiology* 32, pp. 301–18.
- (2001). „The polyvagal theory: phylogenetic substrates of a social nervous system“. In: *International Journal of Psychophysiology* 42.2, pp. 123–146.
- (2003). „Social engagement and attachment: a phylogenetic perspective.“ In: *Ann N Y Acad Sci* 1008, pp. 31–47.
- (2007). „The polyvagal perspective“. In: *Biological Psychology* 74.2, pp. 116–143.
- (2009). „The polyvagal theory: New insights into adaptive reactions of the autonomic nervous system“. In: *Cleveland Clinic Journal of Medicine* 76.Suppl_2, S86–S90.
- (2011). *The polyvagal theory : neurophysiological foundations of emotions, attachment, communication, and self-regulation*. New York: W.W. Norton.
- Proust, Marcel (1919). *A la recherche du temps perdu*. Paris: Gallimard.
- Purdon, Christine (2004). „Empirical investigations of thought suppression in OCD“. In: *Journal of Behavior Therapy and Experimental Psychiatry* 35.2, pp. 121–136.

- Quiroga, Rodrigo Quian and Stefano Panzeri (2009). „Extracting information from neuronal populations: information theory and decoding approaches“. In: *Nature Reviews Neuroscience* 10.3, pp. 173–185.
- Ricoeur, Paul (2003). *La mémoire, l'histoire, l'oubli*. Paris: Seuil.
- Rimé, Bernard (2009). *Le partage social des émotions*. Paris: Presses universitaires de France.
- Roediger, H. L. and K. A. DeSoto (2014). „Forgetting the presidents“. In: *Science* 346.6213, pp. 1106–1109.
- Roediger, Henry L. and Magdalena Abel (2015). „Collective memory: a new arena of cognitive study“. In: *Trends in Cognitive Sciences* 19.7, pp. 359–361.
- Saint-Laurent, Constance de and Sandra Obradović (2018). „Uses of the Past: History as a Resource for the Present“. In: *Integrative Psychological and Behavioral Science*.
- Sakaki, Michiko, Hyun Joo Yoo, Lin Nga, et al. (2016). „Heart rate variability is associated with amygdala functional connectivity with MPFC across younger and older adults“. In: *NeuroImage* 139, pp. 44–52.
- Schacter, D. L. and D. R. Addis (2007). „The cognitive neuroscience of constructive memory: remembering the past and imagining the future“. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1481, pp. 773–786.
- Schacter, D. L., R. L. Buckner, and W. Koutstaal (1998). „Memory consciousness and neuroimaging“. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 353.1377, pp. 1861–1878.
- Schacter, Daniel L., Donna Rose Addis, and Randy L. Buckner (2007). „Remembering the past to imagine the future: the prospective brain“. In: *Nature Reviews Neuroscience* 8.9, pp. 657–661.
- Schie, Kevin van and Michael C. Anderson (2017). „Successfully controlling intrusive memories is harder when control must be sustained“. In: *Memory* 25.9, pp. 1201–1216.
- Schmidt, Helmut, Daniele Avitabile, Ernest Montbrió, and Alex Roxin (2018). „Network mechanisms underlying the role of oscillations in cognitive tasks“. In: *PLOS Computational Biology* 14.9, pp. 1–24.
- Schmitz, Taylor W., Marta M. Correia, Catarina S. Ferreira, Andrew P. Prescott, and Michael C. Anderson (2017). „Hippocampal GABA enables inhibitory control over unwanted thoughts“. In: *Nature Communications* 8.1.
- Sel, Alejandra, Ruben T. Azevedo, and Manos Tsakiris (2016). „Heartfelt Self: Cardio-Visual Integration Affects Self-Face Recognition and Interoceptive Cortical Processing“. In: *Cerebral Cortex*.
- Shaffer, Fred, Rollin McCraty, and Christopher L. Zerr (2014). „A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability“. In: *Frontiers in Psychology* 5.
- Shah, Amit J., Rachel Lampert, Jack Goldberg, et al. (2013). „Posttraumatic Stress Disorder and Impaired Autonomic Modulation in Male Twins“. In: *Biological Psychiatry* 73.11, pp. 1103–1110.
- Shrikanth, Sushmita, Piotr M. Szpunar, and Karl K. Szpunar (2018). „Staying positive in a dystopian future: A novel dissociation between personal and collective cognition.“ In: *Journal of Experimental Psychology: General* 147.8, pp. 1200–1210.
- Skrapari, Ioanna, Nicholas Tentolouris, Despoina Perrea, et al. (2007). „Baroreflex Sensitivity in Obesity: Relationship With Cardiac Autonomic Nervous System Activity“. In: *Obesity* 15.7, pp. 1685–1693.

- Smit, Annika S., Paul A.T.M. Eling, Maria T. Hopman, and Anton M.L. Coenen (2005). „Mental and physical effort affect vigilance differently“. In: *International Journal of Psychophysiology* 57.3, pp. 211–217.
- Smith, Ryan, Julian F. Thayer, Sahib S. Khalsa, and Richard D. Lane (2017). „The hierarchical basis of neurovisceral integration“. In: *Neuroscience & Biobehavioral Reviews* 75, pp. 274–296.
- Speckens, Anne E.M., Ann Hackmann, Anke Ehlers, and Bea Cuthbert (2007). „Imagery special issue: Intrusive images and memories of earlier adverse events in patients with obsessive compulsive disorder“. In: *Journal of Behavior Therapy and Experimental Psychiatry* 38.4, pp. 411–422.
- Staudigl, T., S. Hanslmayr, and K.-H. T. Bauml (2010). „Theta Oscillations Reflect the Dynamics of Interference in Episodic Memory Retrieval“. In: *Journal of Neuroscience* 30.34, pp. 11356–11362.
- Stickgold, Robert and Matthew P Walker (2013). „Sleep-dependent memory triage: evolving generalization through selective processing“. In: *Nature Neuroscience* 16.2, pp. 139–145.
- Stone, Charles B and William Hirst (2014). „(Induced) Forgetting to form a collective memory“. In: *Memory Studies* 7.3. Ed. by Lucas M Bietti, Charles B Stone, and William Hirst, pp. 314–327.
- Stone, Charles B, Theofilos Gkinopoulos, and William Hirst (2017). „Forgetting history: The mnemonic consequences of listening to selective recountings of history“. In: *Memory Studies* 10.3, pp. 286–296.
- Szpunar, Piotr M and Karl K Szpunar (2016). „Collective future thought: Concept function, and implications for collective memory studies“. In: *Memory Studies* 9.4, pp. 376–389.
- Thayer, Julian F and Richard D Lane (2000). „A model of neurovisceral integration in emotion regulation and dysregulation“. In: *Journal of Affective Disorders* 61.3, pp. 201–216.
- Thayer, Julian F., Anita L. Hansen, Evelyn Saus-Rose, and Bjorn Helge Johnsen (2009). „Heart Rate Variability Prefrontal Neural Function, and Cognitive Performance: The Neurovisceral Integration Perspective on Self-regulation, Adaptation, and Health“. In: *Annals of Behavioral Medicine* 37.2, pp. 141–153.
- Troscianko, Emily T (2013). „Cognitive realism and memory in Proust’s madeleine episode“. In: *Memory Studies* 6.4, pp. 437–456.
- Tse, D., R. F. Langston, M. Kakeyama, et al. (2007). „Schemas and Memory Consolidation“. In: *Science* 316.5821, pp. 76–82.
- Tse, D., T. Takeuchi, M. Kakeyama, et al. (2011). „Schema-Dependent Gene Activation and Memory Encoding in Neocortex“. In: *Science* 333.6044, pp. 891–895.
- Tsuji, H, F J Venditti, E S Manders, et al. (1994). „Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study.“ In: *Circulation* 90.2, pp. 878–883.
- Tulving, Endel (1972). „Episodic and Semantic Memory“. In: ed. by Endel Tulving and W. Donaldson, pp. 381–403.
- Tulving, Endel and Donald M. Thomson (1973). „Encoding specificity and retrieval processes in episodic memory.“ In: *Psychological Review* 80.5, pp. 352–373.
- Verguts, Tom (2017). „Binding by Random Bursts: A Computational Model of Cognitive Control“. In: *Journal of Cognitive Neuroscience* 29.6, pp. 1103–1118.
- Viard, Armelle, Gaël Chételat, Karine Lebreton, et al. (2011). „Mental time travel into the past and the future in healthy aged adults: An fMRI study“. In: *Brain and Cognition* 75.1, pp. 1–9.

- Vinik, AI, RE Maser, BD Mitchell, and R Freeman (2003). „Diabetic autonomic neuropathy.“ In: *Diabetes Care* 26, pp. 1553–79.
- Vlasceanu, Madalina, Karalyn Enz, and Alin Coman (2018). „Cognition in a Social Context: A Social-Interactionist Approach to Emergent Phenomena“. In: *Current Directions in Psychological Science*, p. 096372141876989.
- Voltaire (1763/1989). *Traité sur la tolérance*. Paris: GF-Flammarion.
- Wagner, Johanna, Jan R. Wessel, Ayda Ghahremani, and Adam R. Aron (2018). „Establishing a Right Frontal Beta Signature for Stopping Action in Scalp EEG: Implications for Testing Inhibitory Control in Other Task Contexts“. In: *Journal of Cognitive Neuroscience* 30.1, pp. 107–118.
- Waldhauser, G. T., M. Johansson, and S. Hanslmayr (2012). „Alpha/Beta Oscillations Indicate Inhibition of Interfering Visual Memories“. In: *Journal of Neuroscience* 32.6, pp. 1953–1961.
- Waldhauser, Gerd T., Karl-Heinz T. Bäuml, and Simon Hanslmayr (2014). „Brain Oscillations Mediate Successful Suppression of Unwanted Memories“. In: *Cerebral Cortex* 25.11, pp. 4180–4190.
- Waldhauser, Gerd T., Martin J. Dahl, Martina Ruf-Leuschner, et al. (2018). „The neural dynamics of deficient memory control in heavily traumatized refugees“. In: *Scientific Reports* 8.1.
- Wallis, Joni D. (2018). „Decoding Cognitive Processes from Neural Ensembles“. In: *Trends in Cognitive Sciences*.
- Walsh, Matthew M. and John R. Anderson (2012). „Learning from experience: Event-related potential correlates of reward processing neural adaptation, and behavioral choice“. In: *Neuroscience & Biobehavioral Reviews* 36.8, pp. 1870–1884.
- Wegner, Daniel M. (1994). „Ironic processes of mental control.“ In: *Psychological Review* 101.1, pp. 34–52.
- Wegner, DM, DJ Schneider, SR 3rd Carter, and TL White (1987). „Paradoxical effects of thought suppression.“ In: *J Pers Soc Psychol* 53, pp. 5–13.
- Wei, Luqing, Hong Chen, and Guo-Rong Wu (2018). „Structural Covariance of the Prefrontal-Amygdala Pathways Associated with Heart Rate Variability“. In: *Frontiers in Human Neuroscience* 12.
- Wilson, Robert A. (2005). „Collective memory group minds, and the extended mind thesis“. In: *Cognitive Processing* 6.4, pp. 227–236.
- Wimber, Maria, Arjen Alink, Ian Charest, Nikolaus Kriegeskorte, and Michael C Anderson (2015). „Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression“. In: *Nature Neuroscience* 18.4, pp. 582–589.
- Yonelinas, Andrew P. and Maureen Ritchey (2015). „The slow forgetting of emotional episodic memories: an emotional binding account“. In: *Trends in Cognitive Sciences* 19.5, pp. 259–267.

Colophon

This thesis was typeset with $\text{\LaTeX} 2_{\epsilon}$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.

Influences du contrôle inhibiteur et des schémas collectifs sur la mémoire individuelle

Étude en électrophysiologie et imagerie cérébrale fonctionnelle

Résumé

L'oubli est un mécanisme essentiel au fonctionnement de la mémoire. Ceci est vrai lorsque l'on considère la mémoire de l'individu, qui est l'objet privilégié de la neuropsychologie, mais cela est également le cas lorsque nous considérons les usages sociaux, culturels et historiques de la mémoire qui sont les objets d'études privilégiés de l'histoire et des sciences humaines. Dans ce travail, nous avons cherché à comprendre l'influence conjointe des capacités de contrôle inhibiteur et de l'encodage de représentations collectives sur la conservation des souvenirs épisodiques. Nous avons abordé la question de l'oubli volontaire à l'aide du paradigme "*Think/No-Think*" qui permet l'induction expérimentale d'intrusions mnésiques, de leur contrôle et de leur oubli. Nous avons tout d'abord montré, en électrocardiographie et en électroencéphalographie, que les capacités d'oubli volontaire influencent directement la manifestation physiologique des émotions qui peuvent leur être associées. Ce résultat a ensuite été complété par une étude, à l'aide d'analyses de classification multivariées, montrant que les dynamiques neurales qui encadrent la suppression effective des souvenirs, lorsqu'ils apparaissent involontairement à la conscience, sont comparables aux réflexes de réorientation attentionnelle qui peuvent intervenir lors d'interférences sensorielles ou perceptives. Enfin, nous abordons l'influence des schémas présents en mémoire collective sur l'encodage des souvenirs à travers une étude en imagerie par résonance magnétique fonctionnelle.

Mots-clés: oubli, inhibition, régulation émotionnelle, électrocardiographie, électroencéphalographie, schémas mnésiques, mémoire collective, imagerie par résonance magnétique

Abstract

Forgetting is a central mechanism for memory functioning. This is noticeable regarding individual's memory, which is the privileged topic of neuropsychology, but this is still the case when we consider the social, cultural and historical practices of memory which are the privileged topics of history and social sciences. In this work, we tried to understand the joint influence of inhibitory control capabilities and the encoding of collective representations over episodic memories conservation. We addressed the question of voluntary forgetting through the use of the "*Think/No-Think*" paradigm as a way to induce experimental memory intrusions and their control. First, we report with electrocardiography and electroencephalography that voluntary forgetting capabilities directly affects the physiological roots of associated emotions. This result was then complemented, through the use of multivariate decoding, by a study showing that the neural dynamics surrounding the effective suppression of memories, when they involuntarily intrude into consciousness, are comparable to the reorienting reflex occurring during sensory or perceptive interference. Finally, we assess the influence of collective memory schema on the encoding of memories in a magnetic resonance image study.

Keywords: forgetting, inhibition, emotional regulation, electrocardiography, electroencephalography, memory schemas, collective memory, magnetic resonance imaging