

---

**Research Articles: Behavioral/Cognitive**

**Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories**

Pierre Gagnepain<sup>1</sup>, Justin Hulbert<sup>2</sup> and Michael C. Anderson<sup>3,4</sup>

<sup>1</sup>Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France

<sup>2</sup>Bard College, Annandale-on-Hudson, NY 12504, USA

<sup>3</sup>MRC Cognition & Brain Sciences Unit, University of Cambridge, CB27EF, UK

<sup>4</sup>University of Cambridge, Behavioural and Clinical Neuroscience Institute, Cambridge, CB23EB, UK

DOI: 10.1523/JNEUROSCI.2732-16.2017

Received: 26 August 2016

Revised: 12 April 2017

Accepted: 17 April 2017

Published: 30 May 2017

---

**Author contributions:** P.G. analyzed data; P.G., J.H., and M.C.A. wrote the paper; J.H. and M.C.A. designed research; J.H. performed research.

**Conflict of Interest:** The authors declare no competing financial interests.

The authors would like to thank Mike Posner for facilitating this project, and Ean Huddleston, Yalchin Abdullaev, and Scott Watrous for assistance with data collection. This work was supported by UK Medical Research Council grant MC-A060-5PR00 to M.C.A. and United States National Science Foundation grant 0643321 awarded to M.C.A. The authors declare no competing financial interests.

Corresponding author: Dr. Pierre Gagnepain, Inserm-EPHE-UNICAEN U1077, Centre Cyceron, Boulevard Becquerel, BP 5229, F-14074 Caen Cedex, Email: [gagnepain@cyceron.fr](mailto:gagnepain@cyceron.fr); Phone: +33(0)231470159

**Cite as:** J. Neurosci ; 10.1523/JNEUROSCI.2732-16.2017

**Alerts:** Sign up at [www.jneurosci.org/cgi/alerts](http://www.jneurosci.org/cgi/alerts) to receive customized email alerts when the fully formatted version of this article is published.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2017 Gagnepain et al.

# 1 **Parallel Regulation of Memory and Emotion Supports the** 2 **Suppression of Intrusive Memories**

3  
4 Pierre Gagnepain<sup>1</sup>, Justin Hulbert<sup>2</sup> and Michael C. Anderson<sup>3,4</sup>

5  
6 <sup>1</sup> Normandie Univ, UNICAEN, PSL Research University, EPHE, INSERM, U1077, CHU de  
7 Caen, Neuropsychologie et Imagerie de la Mémoire Humaine, 14000 Caen, France

8 <sup>2</sup>Bard College, Annandale-on-Hudson, NY 12504, USA

9 <sup>3</sup>MRC Cognition & Brain Sciences Unit, University of Cambridge, CB27EF, UK

10 <sup>4</sup>University of Cambridge, Behavioural and Clinical Neuroscience Institute, Cambridge,  
11 CB23EB, UK

## 12 13 **Corresponding author:**

14 Dr. Pierre Gagnepain

15 Inserm-EPHE-UNICAEN U1077

16 Centre Cyceron,

17 Boulevard Becquerel, BP 5229, F-14074 Caen Cedex

18 Email: [gagnepain@cyceron.fr](mailto:gagnepain@cyceron.fr); Phone: +33(0)231470159

19  
20 **Abbreviated title:** Parallel suppression of intrusions and affect

21  
22 **Manuscript information:** 28 pages, 7 figures, 5 tables, 218 words in Abstract, 677 words in  
23 Introduction, and 1619 words in Discussion

## 24 25 **Acknowledgments :**

26 The authors would like to thank Mike Posner for facilitating this project, and Ean Huddleston,  
27 Yalchin Abdullaev, and Scott Watrous for assistance with data collection. This work was  
28 supported by UK Medical Research Council grant MC-A060-5PR00 to M.C.A. and United  
29 States National Science Foundation grant 0643321 awarded to M.C.A. The authors declare no  
30 competing financial interests.

34 **Abstract**

35 Intrusive memories often take the form of distressing images that emerge into a person's  
36 awareness, unbidden. A fundamental goal of clinical neuroscience is to understand the  
37 mechanisms allowing people to control these memory intrusions and reduce their emotional  
38 impact. Mnemonic control engages a right fronto-parietal network that interrupts episodic  
39 retrieval by modulating hippocampal activity; less is known, however, about how this  
40 mechanism contributes to affect regulation. Here we report evidence in humans (males and  
41 females) that stopping episodic retrieval to suppress an unpleasant image triggers parallel  
42 inhibition of mnemonic and emotional content. Using functional magnetic resonance imaging,  
43 we found that regulation of both mnemonic and emotional content were driven by a shared  
44 fronto-parietal inhibitory network, and were predicted by a common profile of MTL down-  
45 regulation involving the anterior hippocampus and the amygdala. Critically, effective  
46 connectivity analysis confirmed that reduced amygdala activity was not merely an indirect  
47 consequence of hippocampal suppression; rather, both the hippocampus and the amygdala  
48 were targeted by a top-down inhibitory control signal originating from the dorsolateral  
49 prefrontal cortex. This negative coupling was greater when unwanted memories intruded into  
50 awareness and needed to be purged. Together, these findings support the broad principle that  
51 retrieval suppression is achieved by regulating hippocampal processes in tandem with  
52 domain-specific brain regions involved in reinstating specific content, in an activity-  
53 dependent fashion.

54

55 **Significance statement**

56 Upsetting events sometimes trigger intrusive images that cause distress, and that may  
57 contribute to psychiatric disorders. People often respond to intrusions by suppressing their  
58 retrieval, excluding them from awareness. Here we examined whether suppressing aversive  
59 images might also alter emotional responses to them, and the mechanisms underlying such  
60 changes. We found that the better people were at suppressing intrusions, the more it reduced  
61 their emotional responses to suppressed images. These dual effects on memory and emotion  
62 originated from a common right prefrontal cortical mechanism that down-regulated the  
63 hippocampus and amygdala in parallel. Thus, suppressing intrusions affected emotional  
64 content. Importantly, participants who did not suppress intrusions well showed increased  
65 negative affect, suggesting that suppression deficits render people vulnerable to psychiatric  
66 disorders.

67

68

69

70

71

72

73

74

75

76

## 77 Introduction

78 Sometimes the past intrudes upon the present. Although a passing disturbance for most  
 79 people, such intrusive memories can be vivid, persistent, and distressing for individuals  
 80 suffering from post-traumatic stress, anxiety, or obsessive-compulsive disorders (Brewin et  
 81 al., 2010). Indeed, distressing images are believed to both precipitate psychopathological  
 82 symptoms and contribute to their maintenance (e.g. Brewin et al., 2010; Moritz et al., 2014;  
 83 Rachman, 2007). Understanding why some people have difficulty controlling memories  
 84 requires that we characterize the neural systems that inhibit memory intrusions and that  
 85 attenuate the distress they cause. Here we examine how people suppress the retrieval of  
 86 intrusive images, focusing on whether and how this process contributes to regulating affect.

87 Despite differing goals, memory control and affect regulation engage similar brain regions.  
 88 For example, suppressing retrieval engages the right middle frontal gyrus (MFG) and reduces  
 89 retrieval-related hippocampal activity (Anderson et al., 2004; Depue et al., 2007; see  
 90 Anderson and Hanslmayr, 2014; Anderson et al., 2016, for reviews). Effective connectivity  
 91 analyses indicate that these reductions arise from inhibitory modulation by the MFG (Benoit  
 92 and Anderson, 2012; Gagnepain et al., 2014) that increase forgetting of suppressed traces and  
 93 reduce their tendency to intrude involuntarily (Benoit et al. 2015). Similarly, regulating  
 94 emotional responses to negative stimuli engages right MFG (Hayes et al., 2010; Ochsner et  
 95 al., 2004; Eippert et al., 2007; Kim and Hamann, 2007) to suppress emotion-related activity in  
 96 the amygdala (Comte et al., 2016; Kohn et al., 2014; Radaelli et al., 2014). Comparisons of  
 97 retrieval suppression and affect regulation confirm their overlapping localization within the  
 98 MFG (Depue et al., 2015). Critically, suppressing retrieval of aversive images without affect  
 99 regulation instructions reduces hippocampal and amygdala activity (Depue et al., 2007; Depue  
 100 et al. 2010), suggesting that suppression regulates memory and affect during unpleasant  
 101 intrusions.

102 How retrieval suppression might regulate negative affect is unclear. One possibility is that  
 103 suppression down-regulates hippocampal activity, preventing reinstatement of upsetting  
 104 imagery; this may truncate input into the amygdala, reducing its activity and pre-empting  
 105 emotional responses to the memory. Alternatively, suppression may inhibit both hippocampal  
 106 and amygdala processes, rendering unpleasant memories less intrusive and upsetting. This  
 107 form of parallel modulation has precedent. For example, effective connectivity analyses show  
 108 that MFG inhibits both hippocampal and fusiform cortex activity when participants suppress  
 109 memories of visual objects (Gagnepain et al, 2014); this modulation predicts reduced priming  
 110 in fusiform cortex on later perceptual identification tests, indicating that the objects' sensory  
 111 representations were suppressed. Gagnepain et al. argued that retrieval cues had triggered  
 112 intrusions, driving re-entrant signals from the hippocampus to re-instate the objects' sensory  
 113 features. Suppressing object memories might therefore have engaged inhibitory control  
 114 targeted at both hippocampus and visual cortex. Analogously, suppressing emotional images  
 115 may trigger inhibitory processes targeted at emotion and scene features reinstated during  
 116 intrusions (Figure 1A; Gagnepain et al., 2014).

117 To test this parallel modulation hypothesis, we conducted functional magnetic resonance  
 118 imaging (fMRI) as participants suppressed episodic retrieval. Participants performed the  
 119 Think/No-Think (TNT) task, which included trials requiring them to attend to a reminder (a  
 120 face) of a scene that was either aversive or neutral; for each reminder, they were cued to  
 121 retrieve the scene (Think items), or to suppress its retrieval (No-Think items) (Figure 1B).  
 122 After each trial, participants classified whether the reminder elicited awareness of its paired  
 123 scene (Levy and Anderson, 2012), allowing us to isolate when No-Think trials triggered  
 124 intrusions. After this phase, participants rated the Think and No-Think scenes' valence, along  
 125 with previously studied Baseline scenes not presented during this TNT phase. We tested three

126 predictions using behavioral partial least squares (PLS) (McIntosh and Lobaugh, 2004;  
 127 Krishnan et al., 2011) and dynamic causal modeling (Friston et al., 2003). First, we  
 128 considered whether a common area within right MFG exists whose intrusion-related  
 129 activations predict both better intrusion control and reduced negative affect for suppressed  
 130 scenes. Second, we examined whether shared suppression-related deactivations in the  
 131 hippocampus, parahippocampus, and amygdala predict both intrusion control and reduced  
 132 affect. Third and critically, effective connectivity analyses should reveal that right MFG  
 133 modulates the hippocampus, parahippocampus, and amygdala in parallel, and that intrusions  
 134 trigger greater negative coupling.

## 135 **Materials and Methods**

### 136 *Experimental design*

137 *Participants* - Twenty-four right-handed native English speakers between the ages of 18 to 35  
 138 years were paid to participate (8 males). They had no reported history of neurological,  
 139 medical, visual or memory disorders. The project was approved by both the University of  
 140 Oregon Institutional Review Board and Cambridge Psychology Research Ethics Committee,  
 141 and all participants gave written consent. Participants were asked not to consume  
 142 psychostimulants, drugs, or alcohol prior to the experimental period. Two participants were  
 143 excluded given that they had an insufficient number of intrusions - less than 5% (i.e. 4 trials)  
 144 within a given emotional condition (see below) - for the purposes of fMRI analyses.

145 *Material* - The stimuli were 48 face-scenes pairs plus 10 filler pairs selected from the  
 146 International Affective Picture System (IAPS; Lang et al., 2008) database. Half of the critical  
 147 scenes were normed as Negative (IAPS number: 1301, 2053, 2141, 2700, 2710, 2900, 3280,  
 148 6020, 6244, 6571, 6831, 9041, 9042, 9102, 9180, 9181, 9320, 9420, 9470, 9520, 9561, 9800,  
 149 9830, 9911; Mean valence = 2.8; SD valence = 1.7; Mean arousal = 5.3; SD arousal = 2.3),  
 150 whereas the other half were Neutral (IAPS number: 1121, 1313, 1640, 1810, 2250, 2487,  
 151 2616, 4100, 4535, 5395, 5455, 5628, 7289, 7351, 7402, 7480, 7495, 7503, 7510, 7560, 7640,  
 152 8060, 8117, 8250; Mean valence = 5.9; SD valence = 1.7; Mean arousal = 4.8; SD arousal =  
 153 2.2). Three lists of 8 pairs (assigned to Think, No-Think, and Baseline conditions) were  
 154 created for each valence condition and were counterbalanced so that they appeared in each  
 155 TNT condition equally often, across participants.

156 *Procedure* - To match the strength of initial encoding for all pairs, participants first learned all  
 157 face-scenes pairs through a drop-off/feedback cycle procedure. After studying all pairs for 6  
 158 seconds each, participants were given test trials presenting the face cue for a given pair for up  
 159 to 4 seconds and asked whether they could recall and fully visualize the paired scene. If so,  
 160 three scenes then appeared (one correct and two foils taken from other pairs), and they  
 161 received up to 5 seconds to select which scene went with the face cue. After selecting a scene  
 162 or if the response window expired, a screen appeared for 1 second indicating whether the  
 163 recognition judgment was correct, incorrect, or was not registered before the end of the trial.  
 164 In all cases (even if participants indicated that they could fully visualize the associated scene  
 165 in the first step), each trial ended with the correct pairing appearing onscreen for 3.5 seconds.  
 166 Participants were asked to use this feedback to increase their knowledge of the pair. Once all  
 167 pairs had been presented, all of the face-scene pairs not recalled or correctly recognized were  
 168 presented again in a randomized order until each pairing had been correctly identified once.  
 169 After testing all pairs in this manner, a second drop-off/feedback cycle was applied in the  
 170 same manner, thus ensuring accurate and strong memory for all pictures. This procedure  
 171 ensured that all pairs were learned to a comparable degree and that any encoding advantage  
 172 for negatively valenced stimuli was carefully controlled.

173 Following learning and after practice with the TNT task, participants entered the MRI

174 scanner. At this point, participants engaged in a final round of TNT practice that was followed  
 175 by a brief reminder of all of the studied pairs (1.5 seconds each), during which participants  
 176 were asked once again to reinforce their knowledge of the pairings. This overtraining  
 177 procedure was intended to ensure that images would intrude when its cue was presented  
 178 during the TNT phase, allowing us to isolate brain regions engaged to control these intrusions.

179 Participants then performed the Think/No-Think task, which was divided into 5 sessions, each  
 180 7-8 minutes in length. Each session presented two repetitions of 16 Think (8 Negative and 8  
 181 Neutral face-cues) and 16 No-Think (8 Negative and 8 Neutral face-cues) items, yielding,  
 182 across the 5 sessions, 160 trials per condition in total (32 trials x 5 sessions). Cues appeared  
 183 for 3 seconds either framed in green or red, centered on a black background. On Think trials,  
 184 the cue was bounded by a green box, and participants were told to generate as detailed and  
 185 complete an image of the associated scene as possible. On No-Think trials, the cue was  
 186 bounded by a red box, and participants were told that it was imperative to prevent the scene  
 187 from coming to mind at all and that they should fixate and concentrate on the face-cue  
 188 without looking away (they knew their eye movements were being monitored). During reduced  
 189 trials, participants were asked to block thoughts of the scene by blanking their mind and  
 190 not by replacing the scene with any other thoughts or mental images. If the object image came  
 191 to mind anyway, they were asked to push it out of mind.

192 After the offset of each of the Think or No-Think trial cues, participants reported the extent to  
 193 which the associated scene had entered awareness by pressing one of three buttons  
 194 corresponding to the labels: never, briefly, often. Although participants had up to 10 seconds  
 195 to make this rating, they were instructed and trained to make this rating quickly without  
 196 thinking about the associated picture. Their response was followed by a jittered fixation cross  
 197 lasting 500-8000 ms ( $M = \sim 2200$  ms;  $SD = \sim 2000$  ms depending on sessions), optimized to  
 198 increase the efficiency of the event-related response estimation. These “intrusion ratings”  
 199 were used to isolate trials with intrusive memories and quantify their occurrence. Specifically,  
 200 we used participants’ responses to classify each trial as either having an intrusion (i.e. a  
 201 “briefly” or “frequent” response) or not (a “never” response) in binary fashion. For each  
 202 repetition of a given condition (e.g. the No-Think, negative-valence condition), we averaged  
 203 these binary intrusion reports across all 8 items in that condition to compute an intrusion  
 204 proportion for that repetition. We then averaged these intrusion proportions across the 10  
 205 repetitions of TNT instruction (i.e., across the 2 repetitions in each of the 5 sessions) to derive  
 206 the overall intrusion rate for a given participant. At the end of the 5 TNT sessions, participants  
 207 also performed a spatial cueing task for 8 minutes (data not reported here).

208 Outside the scanner, aftereffects of memory suppression were examined via an associative  
 209 recognition memory test on Think and No-Think items. In addition, a third group of items  
 210 (Baseline items) was tested during this task. These items had been trained in the same initial  
 211 learning session with the Think and No-Think items, but were omitted from the Think/No-  
 212 Think phase. Because these items were trained at the same time as the Think and No-Think  
 213 items, but did not participate in the Think/No-Think task, they provide a baseline estimate of  
 214 memorability of the scenes, given that neither suppression nor retrieval had been performed  
 215 on them. These pairs enabled us to assess the effects of retrieval and suppression on the  
 216 retention of Think and No-Think items respectively, against pairs that were similarly old.  
 217 During trials of this recognition task, a single face cue was presented for 5000 ms along with  
 218 a target and a single foil sampled from other pairs. We expected nearly perfect recognition  
 219 accuracy on this test; thus, we focused on whether repeated suppression of scene memories  
 220 slowed subsequent recognition. Accordingly, participants were instructed to make their  
 221 response as quickly as they could. Following their response, a 4-point confidence scale  
 222 appeared on the screen and participants had to indicate whether they were 1 (not confident) to



223 4 (highly confident) about their response. We chose to measure recognition memory in this  
 224 procedure instead of using recall tests more typically used in retrieval suppression studies, out  
 225 of concern that our overtraining regimen (used to ensure sufficient intrusions) would cause  
 226 ceiling effects in recall. By measuring recognition, we nevertheless could calculate  
 227 recognition speed for all items, providing a way to assess suppression-induced forgetting,  
 228 despite ceiling effects.

229 In the final phase, participants were asked to rate a series of pictures using Self-Assessment-  
 230 Manikin (SAM) pictorial scale (Lang et al., 2008) to measure whether retrieval suppression  
 231 influenced later affective responses to scenes. During this task, we included not only the  
 232 Think and No-Think items, but also the aforementioned Baseline items, to enable us to  
 233 determine whether retrieval or suppression altered the valence of items, relative to Baseline  
 234 pairs. This 9-point scale quantifies participants' self-reported emotional response to each  
 235 visual scene, reported on a scale labeled with cartoon-like manikins with varying emotional  
 236 facial expressions. Participants selected the numbered facial expression most closely matching  
 237 their perception of the valence of the scene. Because the SAM valence scale includes  
 238 "neutral" ratings mid-scale [with negative affect on the low end (score 1-4), and positive  
 239 affect on the high end (scores 6-9)], higher valence scores for suppressed Negative scenes  
 240 likely reflect diminished negative affect (e.g., moving from a 1 to a 3 involves a movement  
 241 towards neutral); in contrast, higher valence scores for neutral scenes likely reflect increasing  
 242 positive affect. Participants were instructed to press a number from 1 (corresponding to a  
 243 frowning face on the far left of the scale) if a picture made them feel completely unhappy to 9  
 244 (corresponding to a smiley face on the far right if a picture made them feel completely happy).  
 245 If participants felt neutral, neither happy nor sad, they were then instructed to press the 5 key  
 246 under the figure in the middle.

247 We used the foregoing ratings to compute valence scores for each condition for each  
 248 participant, adjusted for pre-experimental valence. This adjustment assumes that the SAM  
 249 valence rating that a participant gives for a scene is likely to be influenced by 3 contributing  
 250 components: the scene's initial valence + the effects of our experimental manipulation for that  
 251 participant (here, the effects of the TNT task) + noise (e.g. attentional fluctuation, etc). To  
 252 account for a part of the unexplained variance and provide a better estimation of the impact of  
 253 TNT manipulation, we expressed the SAM valence rating given by a particular subject for  
 254 each item as a percentage relative to that item's IAPS normative value (Lang et al., 2008),  
 255 which gives a sample-based estimate of the probable starting valence for that scene. We  
 256 performed this correction because we lacked pre-experiment measures of scene valence that  
 257 were specific to each participant, so that we could remove variance across conditions  
 258 contributed by pre-manipulation item differences (above and beyond our attempts to match  
 259 this during material selection), attenuating list-based bias. Ideally, future studies should  
 260 compute this adjustment for starting valence using participant-specific ratings of the scenes  
 261 collected prior to the TNT phase, rather than sample-based normative values.

262 Outlier trimming was applied at the item level to recognition reaction times and adjusted  
 263 SAM valence rating data, with outliers defined as greater than 2 times above or below median  
 264 absolute deviation (MAD; Leys et al., 2013) computed separately for each condition. MAD is  
 265 a robust measure of dispersion given by the following formula:  $MAD = b M_i(|x_i - M_j(x_j)|)$ ,  
 266 where  $x_j$  is the  $n$  original observations,  $M_i$  the median, and  $b = 1.4826$ , a constant linked to  
 267 the assumption of normality of the data, disregarding the abnormality induced by outliers  
 268 (Leys et al., 2013). Details of statistical analyses for behavioral data can be found in the  
 269 Result section.

270

271 ***Imaging Acquisition Parameters***

272 Scanning was performed on a 3-T Siemens Allegra MRI system at the Lewis Center for  
 273 Neuroimaging at the University of Oregon using a 32-channel whole-head coil. High-  
 274 resolution (0.5×0.5×1 mm) T1-weighted image was collected for anatomical visualization and  
 275 normalization. Functional data were acquired using a gradient-echo, echo-planar pulse  
 276 sequence (repetition time = 2,000 ms, echo time = 30 ms, 31 interleaved slice acquisition,  
 277 3×3×3 mm voxel size). The first eight volumes of each session were discarded to allow for  
 278 magnetic field stabilization.

279 ***fMRI Preprocessing***

280 Data were analyzed using Statistical Parametric Mapping software (SPM12, Wellcome  
 281 Department of Imaging Neuroscience, London, UK; RRID:SCR\_007037). During pre-  
 282 processing, images were first corrected for slice acquisition temporal delay before being  
 283 spatially realigned to correct for motion. Images were then normalized using the parameters  
 284 derived from the nonlinear normalization of individual gray-matter T1 images to the T1  
 285 template of the Montreal Neurological Institute (MNI) and spatially smoothed using a 10-mm  
 286 FWHM Gaussian kernel for second-level univariate analyses. Note, however, that native  
 287 space 4-mm smoothed images were used for MTL ROI analyses in order to ensure maximum  
 288 accuracy and demarcation between MTL hand-drawn ROIs (i.e., parahippocampal cortex,  
 289 hippocampus, and amygdala).

290 ***ROI definition for activation, brain-behavior correlation, and DCM analyses***

291 We defined an identical set of ROIs across activation, brain-behavior correlation, and  
 292 Dynamic Causal Modeling (DCM; Friston et al., 2003) analyses as follows. Given the fine-  
 293 grain demarcation between the amygdala and the hippocampus, warping during the image  
 294 normalization stage may introduce spatial errors, mixing up voxels from these structures. To  
 295 address this issue, we first made hand-drawn, participant-specific MTL masks, based on the  
 296 individual T1-weighted image. Anatomical demarcation was done according to Franko et al.  
 297 (2014) and Pruessner et al. (2000; 2002).

298 Within these participant-specific native MTL masks, the maximum peak was identified using  
 299 the No-Think < Think contrast. Then, from this peak, an in-house program was then used to  
 300 select the most significant contiguous voxels corresponding to 10% of the total mask size. A  
 301 new mask was created from these voxels with an average volume across participants of 200  
 302 mm<sup>3</sup>, 245 mm<sup>3</sup>, and 160 mm<sup>3</sup> for the parahippocampal cortex, hippocampus, and amygdala,  
 303 respectively. Selecting the voxels around the strongest peak is important for DCM analyses  
 304 because effective connectivity analyses are meaningless in the absence of univariate effects.  
 305 However, in this context, creating a sphere around the peak as is often done would not have  
 306 been appropriate because such spheres could include voxels from different MTL structures.  
 307 Our approach thus ensures that selected voxels exhibit a strong univariate effect between  
 308 Think and No-Think conditions and also respect anatomical boundaries. Within this mask, we  
 309 extracted parameter estimates using non-normalized images (i.e. participant's native space) to  
 310 ensure maximum accuracy and demarcation between ROIs and computed the Intrusion *versus*  
 311 Non-Intrusion contrast (for each valence condition). We note the contrast used to select this  
 312 mask (collapsing across all No-Think trials *versus* Think trials) is orthogonal to the  
 313 comparison of Intrusion *versus* Non-Intrusion trials, thus avoiding circularity issues when we  
 314 compared those conditions. Figure 3B reports the individual peak foci once projected back to  
 315 the normalized MNI space for illustrative purpose.

316 We used a similar procedure to create an ROI that reflects activity relating to control. We  
 317 focused on the anterior section of the right MFG which encompasses the putative supramodal



inhibitory control region spanning motor, memory, and emotion inhibition described in Depue et al. (2015). Given that this putative supramodal region lies in the anterior section of the right MFG, but that there are no clear anatomical boundaries to define it (unlike the hippocampus), we defined an initial binary mask based on the cluster centered alongside the right MFG from the group level No-Think > Think contrast. We then restricted this mask to voxels with “y” coordinates above 30 mm in MNI space (corresponding to the anterior half of the MFG whose coordinates approximately range from 0 to 60 mm on the “y” axis). Note that two local maxima within this functional cluster are close to the supramodal region described in Depue et al. (2015), and were located in the following MNI coordinates: [x = 30; y = 48; z = 16] and [x = 28; y = 48; z = 32] (see Table 2). This MFG mask was then projected back into participants’ native spaces using inversed normalization parameters. Within these participant-specific native masks, the individual peak maximum was identified using the No-Think > Think contrast and the most significant contiguous voxels corresponding to 5% of the total mask size were selected (to account for the bigger initial mask volume compared to MTL mask). A new mask was created from these voxels with an average volume across participants of 460 mm<sup>3</sup>.

#### Statistical analyses of fMRI data

*Think/No-Think Univariate Analyses* - The pre-processed time series in each voxel from the main Think/No-Think task was concatenated across sessions to facilitate subsequent DCM analyses. Regressors within a GLM for each voxel were created by convolving a boxcar function (modeled as a 3-second short-epoch) at stimulus onset for each condition of interest (i.e. Think, Intrusion and Non-Intrusion for both Negative and Neutral scenes) with a canonical HRF. In addition to the regressors of interest, further regressors of no interest were included, specifically the six realignment parameters, sines and cosines of up to three cycles per run to capture low-frequency drifts, and constant terms to remove the mean of each run. Filler items, along with the few items with no button press or not recalled during Think condition, were also entered into a single regressor of no interest. For ROI analyses (see ROI definition section), individual parameter estimates were then extracted and averaged in each ROI. ROI were analyzed using analyses of variance (ANOVAs) with Hemisphere, Region, Emotion, Awareness (Intrusion *versus* Non-Intrusion) as within-subject factors for MTL, and Emotion, Awareness (Intrusion *versus* Non-Intrusion) as within-subject factors for the right MFG. Planned comparisons between experimental conditions of interest were performed using paired t-test. Voxel-based analyses were also performed by entering first-level activation maps for each condition of interest into flexible analyses of variance (ANOVAs) implemented in SPM (RRID:SCR\_007037), which used pooled error and correction for non-sphericity to create t statistics. The SPMs were thresholded for voxels whose statistic exceeded a peak threshold corresponding to  $P < .05$  family-wise error (FWE) correction across the whole brain or within the appropriate search volumes of interest using random field theory.

*Brain-Behavior Correlation Analyses* - To compute brain-behavior relationships, we used a robust statistical approach based on the robust-correlation toolbox (Pernet et al., 2013). First, we rejected the null hypothesis based on the percentile bootstrap Confidence Interval (CI), an approach less sensitive to heteroscedasticity of the data than the traditional *t*-test. Second, we corrected those bootstrapped CI for multiple comparisons across our six MTL ROIs (left and right parahippocampal cortex, hippocampus, and amygdala), yielding a 99.3% CI. Third, we used skipped correlations (accounting for bivariate outliers using the S-estimator deviation rule; Rousseeuw and Van Drissen, 1999) which estimate the true association with accurate false positive control and without loss of power.

For each ROI, we computed the Intrusion versus Non-Intrusion contrast and correlated this

with both intrusion proportion (the number of No-Think trials for which an intrusion was reported divided by the number of No-Think trials across sessions) and affect suppression score (No-think – Baseline adjusted rating).

*Behavioral Partial Least Square (PLS) Correlation Analyses* - ROI-behavior correlation analyses do not, however, allow us to ascertain whether the exact same voxels contribute to both mnemonic and affective regulation. Indeed, the hypothesis of a shared regulation mechanism across domains proposes that a shared set of voxels across the control network produces relationships for affect and memory suppression. However, ROI-behavior correlations could, in principle, be supported by different sets of voxels. Behavioral PLS correlation (McIntosh and Lobaugh, 2004; Krishnan et al., 2011) is ideally suited to disentangle this issue and examine more closely the relationship between neural markers of inhibition and behavioral scores across the putative memory control network (see below) and also MTL voxels. Behavioral PLS is a multivariate technique that reduces a set of voxels (i.e., variables) into a ranked series of independent latent variables (LV) that express the largest possible co-variance (or correlation) with behavioral scores. Put simply, this technique tries to identify separate sets of voxels that express quantitatively different relationships with behavioral measures. Voxel activity first has to be aligned and stacked across participants into a brain activation matrix  $X$  of 22 rows (i.e., participants) and  $N$  voxels. Normalized brain images are therefore used for that purpose. In a first series of PLS analyses, the mask used to create brain activity matrix  $X$  was extracted from the univariate analysis of No-Think > Think contrast ( $P < .001$  uncorrected) and reflected memory control activity (i.e. fronto-parietal control network). Within this mask, the Intrusion > Non-Intrusion contrast (for each type of emotional material) was computed for each voxel and the resulting vector of voxels was then stacked across participants. The same procedure was applied to the neural marker of down-regulation in MTL (i.e., Non-Intrusion > Intrusion contrast). Given that this procedure requires the voxels to be aligned across participants, the MTL mask was derived by combining left and right parahippocampal cortex, hippocampus, and amygdala, defined anatomically using the AAL atlas (Tzourio-Mazoyer et al., 2002; RRID:SCR\_003550). Note that the AAL atlas only includes a demarcation of the entire parahippocampal gyrus which was then divided by splitting the mask into anterior (corresponding to perirhinal cortex) and posterior portions to isolate parahippocampal cortex.

Both intrusion proportion and affect suppression scores were entered into a  $Y$  matrix with participants representing rows.  $Y$  and  $X$  data tables were then mean-centered. Each participant vector of activation was additionally normalized such that the sum of squares of all its voxel values was equal to one. This normalization ensured that the voxels of each participant's vector of activation now have the same variance and that differences between participants are not due to overall differences in activation (Abdi et al., 2012). The cross-block product of  $X$  and  $Y$  (i.e.  $Y^T * X$ ) hence produced a matrix ( $R$ ) encoding the relationship between each voxel and behavioral measurements across participants. A Singular Value Decomposition (SVD) was then applied to this  $R$  correlation matrix, such that  $R$  is decomposed into three matrices  $R = U * \Delta * V^T$ , with  $U$  being the matrix of behavioral saliences (i.e., weight),  $V$  voxel saliences, and  $\Delta$  the amount of cross-table covariance accounted for by each LV. SVD identifies the latent variables which maximize the covariance between voxel activation ( $X$ ) and behavioral measurements ( $Y$ ). Each latent variable in  $V$  contains a spatial pattern depicting the brain regions where the activation shows the strongest relation to our behavioral scores. The brain scores ( $X^T * V$ ) reflects the summary contribution of each participant's expression of a particular latent variable pattern. Correlations between participants' brain scores and behavioral variables thus indicate how each LV optimally represents relations between behavior and brain activity. The statistical significance of LVs was assessed using 5000

permutations in which participants' brain data matrices were randomly reassigned to behavioral measurements with the singular value recomputed each time. The number of times a singular value exceeds the observed singular value yields the probability of significance of original LVs. To compute the significance of voxel salience, we applied bootstrapping with replacement and recomputed SVD for each new bootstrap sample (5000 in total). This procedure yields a bootstrap distribution of voxel saliences which can then be transformed into a Bootstrapped Standard Ratio (BSR) equivalent to z-score (by dividing initial voxel salience by the standard-error of the bootstrapped distribution) to assess the significance of a given voxel (McIntosh and Lobaugh, 2004). Note that this multivariate technique quantifies the relationship between a voxel and a given dimension and is performed in a single analytic step, and thus does not require correction for multiple comparisons across voxels unlike if multiple univariate independent tests had been performed (McIntosh and Lobaugh, 2004).

*Think/No-Think DCM Analyses* - DCM explains changes in regional activity in terms of experimentally defined modulations ("modulatory input") of the connectivity between regions. Here, we used DCM and Bayesian Model Selection (BMS; Penny et al., 2010) to assess a) whether right MFG suppresses MTL substructures during No-Think trials, b) whether this modulation is targeted only at our emotion-related region of interest (i.e., amygdala), only at our memory-related regions of interest (i.e., parahippocampal cortex and hippocampus) or both memory and emotional regions, and c) whether this inhibitory modulation increases in response to the elevated control demand arising during intrusions. DCM entails defining a network of a few ROIs and the forward and backward connections between them. The neural dynamics within this network are based on a set of simple differential equations (the bilinear state equation was used here) relating the activity in each region to a) the activity of other regions via intrinsic connections in the absence of any experimental manipulation, b) experimentally defined extrinsic input (or "driving input") to one or more of the regions, and, most importantly, c) experimentally defined modulations (or "modulatory input") in the connectivity between regions. Changes in the network dynamics are caused by these driving (entering-regions) or modulatory (between-regions) inputs. These neural dynamics are then mapped to the fMRI time series using a biophysical model of the BOLD response. The neural (and hemodynamic) parameters of this DCM are estimated using approximate variational Bayesian techniques to maximize the free-energy bound on the Bayesian model evidence. Here, BMS was used to select the preferred model at the group level, treating the optimal model across participants as a random effect.

Retrieval inhibition was assumed to originate from right MFG (see Introduction and Results); therefore, we focused on the influence of this region over MTL regions within the same hemisphere as done in previous studies analyzing effective connectivity using the TNT paradigm (Gagnepain et al., 2014; Benoit and Anderson, 2012, Benoit et al., 2015; see also Benoit et al., 2016). To ensure that spurious relationships between MTL substructures were not captured by this analysis of effective connectivity, DCM was performed using participants' native space images smoothed at 4mm to maximize demarcation between MTL ROIs (similar to MTL ROIs analysis; see ROI definition section). For each ROI, the first eigenvariate was extracted and adjusted for effects of no interest (which included the six realignment parameters, sines and cosines of up to three cycles per run to capture low-frequency drifts, and constant terms to remove the mean of each run). The main goal of this analysis was to assess whether and how retrieval inhibition originating from right MFG was transmitted to MTL regions.

Thirty-five DCM models were created (for an illustration of the model space, see Figure 7). All models had bidirectional connections between the parahippocampal cortex and hippocampus, the parahippocampal cortex and amygdala, and between the hippocampus and

465 amygdala. A common input source across models driving network activity was defined as  
 466 Think and No-Think trials entering the right MFG and are meant to represent the influence of  
 467 memory control task instructions in the right MFG. These 35 models could be divided into  
 468 two model families. The first family of models (the Modulatory family) divided the model  
 469 space into five sub-groups that differed according to whether the intrinsic connection from the  
 470 right MFG was additionally modulated or not by No-Think items (modeled here as 3-second  
 471 short-epochs separately for Intrusion and Non-Intrusion of each emotion type). In the first  
 472 sub-group, models included a bottom-up modulation during No-Think trials of the connection  
 473 from the MTL to the MFG. In the second and third sub-groups, this modulation was either  
 474 top-down or bidirectional, respectively. In the fourth sub-group, modulation of MTL activity  
 475 during No-Think trials was assumed to be driven by an afferent input originating from a  
 476 source independent of the MFG. Finally, the fifth sub-group of models did not include any  
 477 additional modulation during No-Think trials. The second family of models (the Regulation  
 478 family) encodes the nature of emotional and memory regulation. In the first subgroup of this  
 479 family (Emotion Regulation; first column in Figure 7), neural activity relating to emotion is  
 480 directly targeted in the amygdala and memory suppression is an indirect consequence of  
 481 down-regulating amygdala activity, which is then echoed in other MTL regions. Therefore,  
 482 this partition includes a single model in which the amygdala is targeted by MFG. The second  
 483 subgroup represents models in which memory-related reactivation is directly targeted in the  
 484 hippocampus and/or parahippocampal cortex and in which changes in emotion-related activity  
 485 in the amygdala is an indirect consequence of memory suppression (Memory Regulation;  
 486 middle column of Figure 7). Finally, the third partition is composed of models in which  
 487 memory and emotion sites are targeted in parallel and therefore include a modulation  
 488 combining parahippocampal cortex and/or hippocampus, and the amygdala (Parallel  
 489 Regulation; right column of Figure 7). After estimating all 35 models for each participant, we  
 490 performed the group BMS as implemented in SPM12 (version DCM12 revision 4750;  
 491 RRID:SCR\_007037). This produces the exceedance probability (i.e., the extent to which each  
 492 model is more likely than any other model) and expected posterior probability (i.e., the  
 493 probability of a model generating the observed data). We also tested whether coupling  
 494 parameters differed significantly from zero using 5000 bootstrapping resamplings of the sum  
 495 of intrinsic connections and modulatory parameters (i.e. DCM.A + DCM.B), and applying  
 496 Bonferroni correction across tested parameters. Statistical differences between coupling  
 497 parameters were assessed with an ANOVA using Region, Emotion, and Awareness as within-  
 498 subject factors.

499 The foregoing model space does not model the full set of pathways that are likely involved in  
 500 achieving mnemonic and affective regulation. For instance, other regions that are likely to be  
 501 involved, based on prior studies, include the anterior cingulate cortex, pre-supplementary  
 502 motor area, right inferior frontal gyrus, ventromedial prefrontal cortex, intraparietal cortex  
 503 and supramarginal gyrus (see Anderson et al., 2016 for a review). We omitted these other  
 504 regions for several reasons. First, it is not presently known which intermediate pathways are  
 505 critical for modulating activity in the regions of interest within the MTL. Second, the  
 506 theoretical question to be addressed in the analysis did not concern these pathways, but rather  
 507 whether MFG was involved in causally influencing mnemonic and affective processing in  
 508 parallel, however that influence might be mediated. Third, the ability to estimate model  
 509 parameters effectively is diminished the more complex models become. We therefore adopted  
 510 the simplest model space that still allowed us to address our key theoretical questions about  
 511 mnemonic and affective control. The winning model is therefore not intended to be a full  
 512 anatomical specification of how mnemonic control is achieved, but rather a focused answer to  
 513 a question about causal dynamics.



## 514 RESULTS

### 515 No reliable differences in rate of learning were observed during training.

516 During training, participants took 1.64 and 1.65 trials, on average, during the first dropoff  
 517 training phase, to learn the negative and neutral face-scene associations, respectively. During  
 518 the second dropoff training cycle, participants required 1 trial, on average (for both stimulus  
 519 classes) to show evidence of having learned the associations. The maximum number of  
 520 exposures was on average 3.62 and 3.64 for negative and neutral scenes, respectively, to reach  
 521 the criterion of perfect performance across both cycles. No reliable differences in rate of  
 522 learning were found.

### 523 Success at suppressing intrusive memories and reducing negative affect of unpleasant 524 experiences are related

525 Repeatedly suppressing retrieval of an unwanted memory previously has been shown to  
 526 decrease its tendency to intrude (Levy et al., 2012; Benoit et al., 2015; Hellerstedt et al., 2016;  
 527 van Schie and Anderson, 2017). Replicating this finding, we found that participants' control  
 528 over intrusions improved with practice (Figure 2A). An Emotion x Block ANOVA on  
 529 participants' intrusion reports for No-Think trials revealed a robust reduction in intrusion  
 530 proportion over blocks [ $F(4, 92) = 11.52, P < .001$ ]. Interestingly, repeated suppressions  
 531 reduced intrusions comparably for Negative and Neutral scenes (neither the main effect of  
 532 Emotion, nor its interaction with Block was significant, all  $F_s < 2.16$ ), inconsistent with the  
 533 notion that negative memories are more difficult to control. Indeed, the average intrusion  
 534 proportions were .344 and .386 for Negative and Neutral scenes, respectively (corresponding  
 535 to averages of 28 and 34 intrusions, respectively), indicating that intrusions were numerically  
 536 less frequent for negative memories. It bears emphasis, however, that these findings were  
 537 observed in a context in which people were actively trying to suppress memory retrieval; it is  
 538 possible that Negative memories might be spontaneously retrieved more often when people  
 539 are not trying to suppress retrieval, a situation that was not studied here. The current findings  
 540 thus indicate that when encoding strength is carefully equated (see Procedure), retrieval  
 541 suppression appears to be comparably effective for materials of Negative and Neutral valence.

542 We then asked if participants' ability to gain control over intrusions across blocks was  
 543 accompanied by persisting suppression of emotional content. To measure whether retrieval  
 544 suppression influenced later affective responses to scenes, we examined the adjusted SAM  
 545 valence ratings (see Figure 1A and Method section). We predicted that suppression would  
 546 reduce negative valence for Negative scenes, especially for people who were effective at  
 547 controlling intrusions. This would be reflected in positive affect suppression scores (No-think  
 548 – Baseline rating) because higher scores on this measure would indicate that No-Think items  
 549 had received less negative ratings than had Baseline items. For Neutral scenes, we did not  
 550 have specific predictions as to whether retrieval suppression should further remove residual  
 551 negativity (which would actually increase positive feelings toward the Neutral scene), or,  
 552 instead, simply dampen positive affect for the item.

553 We first tested whether suppression affected participants' valence ratings for scenes without  
 554 considering how well participants controlled intrusions. This overall analysis revealed no  
 555 differences in valence ratings across Think, No-Think, and Baseline conditions for either  
 556 Negative and Neutral scenes (all relevant comparisons  $P > .1$ ; see Table 1). Thus, across all  
 557 participants, suppression did not consistently affect the perceived valence of the scenes.

558 Whether suppression affects an image's perceived valence would be expected to depend,  
 559 however, on how effectively a person mitigates unwanted intrusions through inhibitory  
 560 control. To test this possibility, we examined the relationship between intrusion frequency and



561 affect suppression scores (No-Think *versus* Baseline). We first correlated participants'  
 562 reported intrusion proportion during No-Think trials with their affect suppression scores. We  
 563 found that participants who were more effective at preventing intrusions of Negative scenes  
 564 during No-Think trials showed greater affect suppression scores ( $r_{\text{skipped}} = -0.46$ ,  $[-0.76 -$   
 565  $0.12]$  bootstrapped 95% CI). We observed no significant relationship between intrusion  
 566 proportion and affect suppression score for Neutral scenes ( $r_{\text{skipped}} = -0.09$ ,  $[-0.35 \ 0.49]$   
 567 bootstrapped 95% CI).

568 To further illustrate this relationship, we split participants into two groups according to how  
 569 well they suppressed intrusions during No-Think trials (Figure 2B). Separating groups in this  
 570 manner qualitatively illustrates the point made by our significant correlation, that participants  
 571 differing in putative memory control skill had differing success in reducing negative affect for  
 572 suppressed scenes. In the high-memory-control group (i.e., low intrusion frequency), we  
 573 observed significantly reduced negative valence for suppressed Negative scenes compared to  
 574 Baseline Negative scenes [ $t(10) = 2.29$ ,  $P < .05$ ]; we observed no evidence that suppression  
 575 altered the valence of Neutral scenes compared to Baseline Neutral scenes ( $t(10) = 0.3$ ,  $P > .1$ ).  
 576 In striking contrast, in the low memory-control group (higher intrusions), Negative scenes that  
 577 participants suppressed during No-Think trials were judged as significantly more negative  
 578 than their Baseline counterparts [ $t(10) = -3.41$ ,  $P < .01$ ], and again, we found no effect of  
 579 suppression on affect for Neutral scenes [ $t(10) = -0.4$ ,  $P > .1$ ]. These relationships between  
 580 mnemonic awareness and changes in perceived valence were specific to No-Think items: no  
 581 relationship between mnemonic awareness and valence effects for Think items was observed,  
 582 regardless of the nature of the scenes (Negative or Neutral).

583 Although the foregoing analysis suggests that people who are successful at suppressing  
 584 aversive intrusions also are better at reducing negative affect, an alternative possibility exists.  
 585 Perhaps the participants who had fewer intrusions in our No-Think task were people who  
 586 simply found the particular No-Think items they received less upsetting during the encoding  
 587 phase. If so, low-intrusion participants may, on our final affect measure, show less negative  
 588 affect for those particular No-Think items than for Baseline items, due to the differing initial  
 589 affective response to those items, not affect suppression. To scrutinize whether these putative  
 590 item effects could account for the relationship between intrusion control and affect  
 591 suppression, we used separate No-Think items to quantify intrusion frequency and affect  
 592 suppression. First, we randomly split No-Think items into a “sample set” and a “test set.”  
 593 Second, we computed the intrusion proportion for a given participant using their “sample set”  
 594 and their affect suppression score using the separate “test set,” rendering these two  
 595 measurements item-independent. Third, we computed the skipped-correlation between these  
 596 intrusion proportion and affect suppression scores. We repeated this three-step process one  
 597 thousand times, randomly splitting No-Think items at each iteration to identify sample and  
 598 test sets. Thus, this procedure yielded a correlation distribution whose reliability could be  
 599 assessed by computing confidence intervals. If the correlation between intrusion frequency  
 600 and affect suppression arose simply because less upsetting items will also be less intrusive,  
 601 then using separate items to quantify these outcomes should eliminate the relationship. This  
 602 did not occur, however: for Negative scenes, we found a significant relationship between  
 603 intrusion proportion and affect suppression that was independent of item selection (mean  $r_{\text{skipped}} = -0.29$ ,  $[-0.51 \ -0.03]$  95% CI), whereas this relation was absent for Neutral scenes  
 604 (mean  $r_{\text{skipped}} = -0.07$ ,  $[-0.30 \ 0.20]$  95% CI). These findings support the possibility that  
 605 better intrusion control truly does relate to superior affect suppression (a possibility  
 606 corroborated by later analyses relating individual differences in brain activity to these scores).  
 607

608 Although the foregoing evidence suggests that suppression disrupts negative affect of No-  
 609 Think scenes (for people with good memory control), another interpretation exists. Instead of

610 disrupting the negative affective content of No-Think items, suppression may instead increase  
 611 the positive valence associated with those suppressed items. If so, we would expect a net  
 612 change in the affect rating for Negative No-Think items towards a more neutral rating,  
 613 reflecting the new mixture of positive and negative valences of those items. One effective way  
 614 to rule out this interpretation would be to include positively valenced items in future retrieval  
 615 suppression studies. If suppression adds positive valence, then positive scenes should be  
 616 rendered even more positive after suppression; but if suppression disrupts valenced content,  
 617 suppression should reduce positive affect for suppressed scenes. Although this alternative  
 618 cannot be ruled out definitively in the current study, data from affective ratings of our Neutral  
 619 Scenes make it less likely. If suppression added positive valence, we would expect to find  
 620 the same correlation between our affect suppression score (No-Think – Baseline) and  
 621 intrusion control for Neutral scenes, but no relationship was found. This finding is more  
 622 consistent with a disruption of affective content, which is only likely to be present in our  
 623 Negatively valenced scenes. Nevertheless, because we did not include positively valenced  
 624 scenes in the current design, it is prudent to limit the current conclusions to the disruption of  
 625 negative affect, not affect more generally.

626 We also examined how suppression affected performance on the associative recognition test  
 627 given at the end of the experiment. As expected, accuracy was near to ceiling, given the  
 628 extensive training of the pairs and the use of a recognition test (accuracy > 95% in all  
 629 conditions). Nevertheless, recognition reaction times revealed aftereffects of suppressing No-  
 630 Think pictures. Averaging across valence types, we observed a significant suppression effect  
 631 for No-Think ( $M = 1464$  ms / $SD = 376$ ) compared to baseline items ( $M = 1393$  ms / $SD = 367$ ;  
 632  $t(21) = 1.89$ ,  $P < .05$ ). Considered separately, suppression effects arose for negative scenes  
 633 (mean suppression effect = 131 ms/ $SD = 288$ ;  $t(21) = 1.82$ ,  $P < .05$ ) but not for Neutral scenes  
 634 (mean suppression effect = 12 ms/ $SD = 132$ ;  $t(21) = 0.54$ ,  $P > 0.1$ ), although this interaction  
 635 did not reach significance (Negative *versus* Neutral), ( $t(21) = 1.42$ ,  $P = .084$ ). These reaction  
 636 time patterns resemble the recall patterns observed by Depue et al (2006) who also studied  
 637 retrieval suppression using the current face-scene pairings (Depue et al., 2006). We also  
 638 observed a significant main effect of valence on recognition time, with Negative Scenes  
 639 showing slower recognition than Neutral scenes ( $t(21) = 5.06$ ,  $P < .001$ ), a pattern observed in  
 640 prior studies examining recognition of scenes and faces (Keightley et al., 2011).

641 Overall, the foregoing findings are consistent with the possibility that suppressing Negative  
 642 scenes reduces their tendency to intrude and may also alter the emotional quality of those  
 643 memories so that their reappearance triggers less negative affect - at least for people who are  
 644 proficient in controlling their memories. Interestingly changes in affect do not arise for  
 645 Neutral scenes suggesting that suppressing unpleasant memories may entail additional  
 646 inhibitory effects not present for Neutral memories. In no case did retrieval during Think  
 647 trials, by contrast, measurably alter the perceived affect of the scenes. More generally, these  
 648 findings suggest that inhibitory control might, in parallel, modulate traces in different  
 649 representational domains (memorial and affective) during efforts to exclude an unwanted  
 650 memory from awareness. Given this possibility, our goal was then to understand how the right  
 651 fronto-parietal control network contributes to this hypothesized parallel regulation of memory  
 652 and emotion, and to determine whether these contributions could be related to an inhibitory  
 653 signal targeted at different substructures within the MTL.

#### 654 **Fronto-MTL regions are engaged during the suppression of intrusive memories.**

655 Before addressing whether retrieval suppression modulates regions related to both memory  
 656 and affect, we confirmed the engagement of the right fronto-parietal control network and the  
 657 disengagement of MTL during retrieval suppression. First, we contrasted No-Think and Think  
 658 trials aggregating conditions over both Negative and Neutral scenes ( $P_{FWE} < .05$ ; see Figure

2C). Consistent with previous findings, we observed more activation during No-Think trials in a large right-lateralized network, including the middle frontal gyrus (MFG), dorsal portion of the anterior cingulate cortex (dACC), superior frontal medial gyrus (SFMG), inferior frontal gyrus (IFG), insula, superior frontal gyrus (SFG), and inferior parietal cortex (IPC). Critically, a right dorsolateral prefrontal cortex cluster, centered on the anterior MFG and previously noted for its involvement in direct memory suppression, survived whole-brain voxelwise correction ( $x = 30$ ,  $y = 48$ ,  $z = 16$ ;  $z_{max} = 6.15$ ; 1591 voxels;  $P_{FWE} < .001$ ; for complete whole-brain analyses, see Table 2). Thus, in agreement with previous findings using other types of suppressed material (e.g., words, objects, faces, places), the right MFG was engaged in the suppression of both Negative and Neutral scenes (Depue et al., 2007; Benoit et al. 2015). Note that we observed no interaction surviving whole-brain correction between memory control (i.e. No-Think > Think) and Emotion, nor did we observe differences between Negative and Neutral scenes in the No-Think condition. These results suggest that a common control network suppresses retrieval irrespective of the valence of the unwanted memory.

We then examined how retrieval suppression affected activity in the right MFG ROI (see ROI selection section). We focused our analysis on No-Think trials, dividing them into Intrusions and Non-Intrusions, based on what participants reported on the intrusion scale immediately following each trial. We note the contrast used to select this ROI (collapsing across all No-Think trials *versus* Think trials) is orthogonal to the comparison of Intrusion *versus* Non-Intrusion trials (up-regulation effect), thus avoiding circularity issues when we compared those conditions. A two-factor ANOVA with Emotion (negative, neutral) and Awareness (Intrusions *versus* Non-Intrusions) as factors revealed neither a main effect of Awareness nor the presence of a significant interaction between the two factors (all  $F_s < 0.94$ ).

We then examined how retrieval suppression affected activity in MTL ROIs. Previous studies have found more pronounced down-regulation of hippocampal activity during retrieval suppression when memories involuntarily intrude into consciousness compared to when they do not (Levy and Anderson, 2012; Benoit et al., 2015). We next sought to replicate these findings to determine whether they generalize to the amygdala, and to examine whether they interact with valence. In agreement with prior findings, a Hemisphere x Region x Emotion x Awareness (Intrusion *versus* Non-Intrusion) ANOVA showed a significant decrease in activity during Intrusion relative to Non-Intrusion trials [ $F(1,21) = 8.79$ ,  $P < .01$ ]. This down-regulation of activity during intrusions varied with both Region and Emotion, as revealed by a significant three-way interaction of Awareness with Region and Emotion [ $F(1,42) = 3.25$ ,  $P < .05$ ]. This interaction was driven mainly by two effects (Figure 3): first, in both hemispheres, the amygdala showed greater down-regulation for negative than for neutral scenes; and second, the left hippocampus and right parahippocampus showed greater down-regulation than the other regions.

Together with engagement of the right fronto-parietal network, these findings support the possibility that retrieval suppression recruited the putative memory inhibition network to suppress episodic retrieval of intrusive memories. Critically, they further show that previous demonstrations of intrusion-dependent down-regulation (Levy and Anderson, 2012) also generalize to the amygdala when aversive scenes are suppressed. This clear evidence for modulation of the amygdala (together with clear negative valence ratings for Negative items [ $M = 2.2$ ] on our final valence rating test), suggests that our training regimen, involving multiple presentations and tests of our face-scene pairs, did not eliminate the affective properties of our negative scenes or the need to suppress emotional responses to them. Importantly, this joint modulation of the amygdala and hippocampus in response to intrusive memories provides strong preliminary support for the parallel regulation of emotion and

memory by retrieval suppression. We explore this parallel regulation hypothesis in greater depth next.

# **Intrusion regulation and affect suppression score depend on common prefrontal regions and MTL down-regulation**

We next examined whether people's ability to suppress intrusive memories and alter their perceived valence (i.e., affect suppression score) might depend on a common right fronto-parietal network that modulates MTL activity. We first used the standard brain-behavior correlation method using a robust statistical approach as described by Pernet et al. (2013; see Method). The down-regulation and up-regulation effects (Intrusion *versus* Non-Intrusion) were extracted in MTL and right MFG ROIs, respectively. We correlated these neural markers of memory suppression with the behavioral markers of affect regulation and intrusion control. Figure 4 reports in detail the outcome of these correlations after correcting for multiple comparisons. We focus on the skipped-correlation which accurately controls for outliers. Broadly, during the suppression of intrusive scenes, both down-regulated activation in the hippocampus and amygdala and up-regulation in right MFG were associated with reduced intrusion frequency for both Negative and Neutral scenes and increased affect suppression scores only for Negative scenes.

The foregoing ROI-behavior correlation analyses provide initial support for a shared control mechanism underlying mnemonic and affective regulation. However, this approach does not allow us to ascertain whether the exact same voxels contribute to regulation in both domains. Indeed, it may be that different sets of voxels across the MTL or the control network produce relationships for affect and intrusion suppression. We therefore used the behavioral PLS correlation method which identify sets of voxels (i.e. dimensions) that express quantitatively different relationships with behavioral measures (see Methods). Behavioral PLS takes the behavioral measurements (here, our intrusion frequency and affect suppression measures) together with fMRI activations observed at each voxel (based here on the Intrusion *versus* Non-Intrusion contrast, our neural marker of intrusion control), and calculates a correlation matrix, across participants. Singular value decomposition is then applied to this correlation matrix to produce orthogonal latent variables (LVs) that optimally represent the relationship between intrusion control activations and behavioral markers of mnemonic and affective control. These LVs are extracted in the order of the amount of covariance they explain between intrusion control activations and behavioral measures. On one hand, if different sets of voxels express a distinct and significant relationships with affect suppression scores and intrusion proportions, respectively, this technique would identify two significantly separate LVs. On the other hand, if a common set of voxels express a common relationship with both affect suppression score and intrusion proportion, a unique LV would emerge from this analysis.

Because the outcome of the PLS analysis might vary with the valence of our materials, we conducted separate analyses for Negative and Neutral scenes. For Negative scenes, the first LV accounted for 76 % and 79 % of the covariance between intrusion control activations and behavioral measures of intrusion proportion and affect suppression within the fronto-parietal retrieval suppression network and the MTL mask, respectively. The first LV was significantly different from random noise as assessed by permutation testing ( $P < .05$ ). For Neutral scenes, the first LV accounted for 73% and 68% of the covariance between behavioral measurements and the prefrontal retrieval suppression network and the MTL mask, respectively. A second LV was not significant for either Negative and Neutral scenes.

The existence of a significant LV that accounts for a large percentage of the covariance between intrusion-control activations and both of our behavioral measures indicates that some



voxels within the prefrontal cortex and MTL are related to mnemonic and/or affective control of unpleasant memories. They do not, however, specify how brain activation relates to those measures. To allow for more clear interpretation of the role of these voxels, we next computed “voxel salience” (the voxels that most contributed to our key LV) and a brain score for each participant (see Method). A brain score indicates how much a given participant expresses the multivariate spatial pattern of correlation between intrusion control activations and behavioral measures of mnemonic and affective control captured by a LV. Thus, correlations between brain scores and behavioral measurements help to identify the direction and the strength of the relationship captured by a given LV (and thus the corresponding voxel salience over that LV).

Critically, within the fronto-parietal control network, we found that for negative scenes, participants’ brain scores for the first LV correlated positively with their affect suppression scores ( $r = 0.60$ ,  $P < .05$ , [0.35 0.79] bootstrapped 95% CI) and negatively with intrusion proportion ( $r = -0.79$ ,  $P < .05$ , [-0.91 -0.60] bootstrapped 95% CI; Figure 5A). Importantly, this finding indicates that for those voxels having a positive salience, up-regulation during Intrusions (*versus* Non-intrusions) negatively correlated with intrusion frequency (i.e., better mnemonic suppression), and additionally positively correlated with affect suppression score (i.e., better affective suppression) for Negative scenes (see Figure 5B for an illustration of this pattern). Note that these patterns are inverted for voxels associated with a negative salience. Voxels within our retrieval suppression mask associated with a significantly positive salience (using a bootstrapping procedure; see Methods), were localized across the entire control network, including left and right MFG, dACC, SFMG, IFG, insula, and IPC (see Figure 5A and Table 3). For Neutral scenes, participants’ brain scores for the first LV also correlated negatively with intrusion proportion ( $r = -0.72$ ,  $P < .05$ , [-0.87 -0.54] bootstrapped 95% CI) but, in contrast to findings with negative scenes, showed no significant correlation with their affect suppression scores ( $r = -0.16$ , [-0.54 0.19] bootstrapped 95% CI; Figure 5A).

Within the MTL mask, we found a similar pattern: for Negative scenes, participants’ brain scores for the first LV correlated positively with affect suppression score ( $r = 0.59$ ,  $P < .05$ , [0.38 0.79] bootstrapped 95% CI) and negatively with intrusion proportion ( $r = -0.55$ ,  $P < .05$ , [-0.78 -0.28] bootstrapped 95% CI), whereas for Neutral scenes, brain scores only correlated with intrusion proportion ( $r = -0.40$ ,  $P < .05$ , [-0.66 -0.02] bootstrapped 95% CI; Figure 6A) but not with affect suppression ( $r = 0.24$ , [-0.12 0.64] bootstrapped 95% CI; Figure 5A) (see Figure 6A). Given that most voxels had a reliably negative salience within the MTL mask (located mainly in the bilateral anterior hippocampus and amygdala; Figure 6A), the foregoing pattern means that during the suppression of intrusive scenes, down-regulating activation in these voxels was associated with decreased intrusion frequency for Negative and Neutral scenes and greater affect suppression scores (but only for Negative scenes; see Figure 6B for an illustration of this pattern). No voxels within the MTL were associated with a significant positive salience (i.e., the opposite pattern).

Taken together, the foregoing results strongly support the hypothesis that retrieval suppression doesn’t simply reduce the accessibility of episodic memories but can also alter negative affect. In particular, increased activation in the right lateral PFC region previously associated with retrieval suppression predicted reduced intrusion frequency as well as increased valence-suppression for those suppressed images. Thus, individuals who adaptively up-regulated right MFG in response to intrusions could purge the unwanted trace from awareness and disrupt emotional content of suppressed traces at the same time, decreasing the likelihood of those traces reentering consciousness and triggering upsetting thoughts. Notably, this evidence for co-localization of mnemonic and affective regulation is unlikely to be an artifact of poor spatial resolution masking what might be two similarly located but distinct forms of control: Because we related activations during retrieval-suppression to mnemonic



805 and affective measures on a voxel-by-voxel basis, these findings likely indicate the operation  
 806 of a shared mechanism. These findings add force to the results of previous within-subjects  
 807 conjunction analyses performed on separate affect regulation and retrieval suppression tasks  
 808 (Depue et al. 2015) and specifically link activations in right MFG to intrusion control and  
 809 affect regulation of negative memories.

810 Critically, our findings further reveal that the ability to control intrusive memories and reduce  
 811 the negative affect associated with them is related to down-regulation of a shared set of voxels  
 812 within the MTL. These shared modulatory targets were localized to anterior hippocampus and  
 813 amygdala. There is evidence for greater anatomical and functional connectivity between  
 814 anterior hippocampus and amygdala (see Poppenk et al., 2013), relative to mid or posterior  
 815 hippocampus supporting this type of interaction. One interpretation of our findings, therefore,  
 816 is that cue-driven reinstatement of aversive scenes in the hippocampus may have triggered  
 817 affect-related activity in the amygdala, prompting the targeting of that structure for inhibitory  
 818 control.

#### 819 **Modulation of MTL activity is parallel and reactive**

820 Although the current findings suggest that suppressing intrusive memories disrupts episodic  
 821 and emotional traces, it does not establish specifically that right MFG modulates MTL  
 822 activity, nor whether such modulation acts in parallel on both memory and emotion-related  
 823 sites. It is possible, for example, that MTL regions were simply not engaged during intrusions,  
 824 rather than being suppressed; participants might have shifted attention away from the  
 825 reminder cue if they sensed that they could not control a memory. Such perceptual avoidance  
 826 might limit sensory input into the MTL, reducing activity during intrusions relative to non-  
 827 intrusions. Moreover, even if right MFG modulates MTL activity, we don't know which  
 828 structures are affected. On the one hand, reduced amygdala activation during intrusions might  
 829 reflect reduced afferent input from memory-related regions (i.e., hippocampus and/or  
 830 parahippocampal cortex), caused by suppressing the latter regions. By this memory-regulation  
 831 account, reduced amygdala is the downstream effect of inhibitory modulation elsewhere. On  
 832 the other hand, modulation instead may target the amygdala and be echoed in memory-related  
 833 regions. Our parallel regulation hypothesis, however, predicts that neither mnemonic nor  
 834 affective regulation alone accounts for the neural data; rather, it states that a top-down  
 835 mechanism regulates emotion-related processing in the amygdala in addition to regulating  
 836 retrieval processes supported by the hippocampal and parahippocampal cortices. Such parallel  
 837 regulation should be particularly pronounced when memories intrude into awareness and  
 838 control must be engaged reactively to counter unwanted awareness.

839 We first used DCM and BMS to determine whether the right MFG modulated MTL activity  
 840 during No-Think trials and, if so, which structures it affected: the right (ipsilateral) amygdala,  
 841 hippocampus, and/or parahippocampal cortex (see Methods). We chose DCM as a method in  
 842 part because any putative causal influence that MFG might exert on activity in the  
 843 hippocampus and amygdala would necessarily be mediated polysynaptically (e.g., perhaps via  
 844 structures such as the anterior cingulate cortex - see Anderson et al., 2016). Because the  
 845 precise anatomical pathway mediating this top-down influence has not yet been identified,  
 846 and because DCM can evaluate effective connectivity without presupposing direct anatomical  
 847 connections between two structures, this method is well suited to test our mechanistic  
 848 hypotheses - hypotheses which mainly concern whether retrieval suppression engages MFG to  
 849 achieve both mnemonic and affective regulation in parallel.

850 To evaluate which regions are modulated during suppression, we compared models lacking  
 851 any modulation (i.e., "null" models), to those characterized by top-down (MFG→MTL),  
 852 bottom-up (MTL→MFG), bidirectional (i.e. top-down and bottom-up), or afferent sources of

modulation (Figure 7). We intended our “afferent” models to capture the possibility that MTL activation differences simply reflected variations in bottom-up sensory input, without modulation of connectivity between MFG and MTL. Top-down or bottom-up modulation models hypothesized that activation differences across Intrusion and Non-Intrusion conditions were driven by modulatory influences on MFG→MTL or MTL→MFG connections, respectively. The findings were unambiguous: BMS overwhelmingly favored models including top-down modulation, with an exceedance probability (EP) = 0.999 and expected posterior probability (EPP) = .80; in contrast, “null”, “bottom-up”, and “afferent” models had a EP and a EPP of 0.0/0.0/0.001, and 0.039/0.039/0.12, respectively. Exceedance probability refers to the extent to which a model is more likely in relation to other models considered, whereas expected posterior probability is the probability of a model generating the observed data.

Having evidence for modulation, we then used BMS to compare the 14 remaining models within the “top-down” sub-family (including top-down and bidirectional models). We divided these into sub-families to distinguish among our three alternatives: the emotion regulation family (models with modulation of the amygdala), the memory regulation family (models with modulation of hippocampus, the parahippocampal cortex, or both), and the parallel regulation family (models with modulation of the amygdala and the hippocampus, the amygdala and the parahippocampal cortex, or all three structures; Figure 7). Critically, in comparing these three families, BMS provided strong evidence favoring parallel regulation models (EP = 0.99, and EEP = 0.63) over either emotion regulation (EP = 0.003, and EEP = 0.16) or memory regulation models (EP = 0.009, and EEP = 0.20). These data thus strongly support our key hypothesis that retrieval suppression drives modulatory signals from the right MFG to MTL regions (together with a bottom-up influence), and that suppression not only affects episodic retrieval processes, but also affective processes.

We also sought to determine whether the greater down-regulation of BOLD signal in the MTL during intrusions reflected stronger action of inhibitory control. To evaluate this hypothesis, we extracted DCM coupling parameters (i.e., DCM.B matrix) and applied Bayesian model averaging (BMA) on the preferred Parallel family. From this model family, we extracted twelve parameters for each of our participants, that quantified the modulatory influence of MFG on the different target regions (i.e., PhC/Hip/Amg), for materials differing in valence (i.e., Negative/Neutral), during differing levels of intrusiveness (i.e., Intrusion/Non-Intrusion) that we measured in our design. We first tested whether these parameters differed significantly from zero using 5000 bootstrapping resamplings of the sum of intrinsic connections and modulatory parameters (i.e. DCM.A + DCM.B), and applying Bonferroni correction across the 12 parameters (leading to 99.6% CI). Note that one participant with aberrant coupling parameters deviating more than 4 standard deviations of the mean was excluded from this analysis. For negative scenes, memory intrusions were associated with significant negative coupling to the parahippocampal cortex ([-0.97 -0.17] bootstrapped 99.6% CI), to the hippocampus ([-0.68 -0.04] bootstrapped 99.6% CI) and to the amygdala ([-0.87 -0.08] bootstrapped 99.6% CI; see Table 5). For intrusions of Neutral scenes, coupling parameters to the parahippocampal cortex ([-1.06 -0.14] bootstrapped 99.6% CI), and to the hippocampus ([-0.70 -0.05] bootstrapped 99.6% CI) did also differ significantly from zero, but not for the amygdala ([-0.70 0.51] bootstrapped 99.6% CI; see Table 5). Modulatory parameters never differed reliably from zero during non-intrusion trials regardless of valence. These findings suggest that negative coupling between the MFG and MTL regions was generally greater during Intrusions than Non-Intrusions, and spanned both memory and emotion-related regions for negative materials.

To verify the impression that top-down coupling parameters differed between Intrusion and

Non-Intrusion trials, we computed a Region (PhC/Hip/Amg) x Emotion (Negative/Neutral) x Awareness (Intrusion/Non-Intrusion) ANOVA. As predicted, we observed a significant main effect of Awareness on coupling parameters which were, on average, more negative during Intrusions ( $M = -0.39$ ;  $SD = 0.38$ ) than during Non-Intrusions ( $M = -0.24$ ;  $SD = 0.45$ ),  $F(1,20) = 3.82$ ,  $P < .05$ . No further main effects or interactions were significant (all  $F_s < 0.92$ ), except for the Region x Emotion x Awareness interaction which approached significance [ $F(1,20) = 2.87$ ,  $P = .068$ ]. This latter trend in part reflects greater negative coupling to the amygdala during intrusions of negative, compared to neutral scenes (see Table 5).

Taken together then, the results of our BMS and BMA analyses indicate the existence of modulatory influences of MFG on MTL structures and that such modulation is inhibitory in nature. During the suppression of negative memories, this modulation not only affects regions critical to episodic memory, but also the amygdala, and is particularly pronounced when memories intrude into awareness and need to be purged. These findings support the parallel regulation of memory and emotion by inhibitory control mechanisms that are reactive in nature, suppressing awareness of intrusive memories.

## Discussion

When unpleasant memories intrude into awareness, people often suppress their retrieval to regulate their emotional state. Although considerable work has addressed the mechanisms of retrieval suppression, this work has not examined how suppression alters people's emotional state. Does suppression only target episodic memories, disrupting mnemonic awareness, reducing input to mechanisms that would have driven unpleasant emotions? Or does it inhibit both episodic memories and affective traces? The present data supports the latter view: suppressing unpleasant reminders not only disrupts memories supported by the hippocampus and parahippocampus, but also emotional traces that depend on the amygdala, and these parallel effects arise from a shared inhibitory mechanism mediated by the right dorsolateral prefrontal cortex.

**Parallel regulation of memory and affect.** Several key findings point to the parallel regulation of memory and affect. First, behavioral and neural data suggest that suppression reduced affective responses to unpleasant memories. Behaviorally, participants who controlled intrusions well showed greater reductions in negative affect for the suppressed unpleasant scenes. This suggests that suppressing episodic retrieval engages a mechanism that also affects emotional traces. Supporting this interpretation, a PLS analysis revealed common areas within the right dorsolateral and ventrolateral prefrontal cortex that predicted both how well participants reduced intrusions and negative affect for suppressed scenes. Intrusions increased activation in these regions, suggesting a co-localized control mechanism that suppresses both types of content. Critically, PLS analyses also identified regions within the anterior hippocampus and amygdala that predicted both intrusion control and affect regulation. Unlike in the prefrontal cortex, voxels in these regions were down-regulated, suggesting that inhibitory mechanisms suppressed their activity. Together, these behavioral and neural findings suggest that suppression mechanisms triggered by intrusions not only disrupt episodic memories (Levy and Anderson, 2012) but also blunt negative feelings about images in an enduring way, by a shared mechanism supported by lateral prefrontal cortex.

Of course, suppression-related activations do not necessarily imply that the prefrontal cortex reduced MTL activation. Reduced activations may instead arise from other sources of modulation, or indeed, may not even reflect active modulation. Contrary to these possibilities, however, DCM analyses robustly confirmed that the right MFG modulates the MTL during suppression (an influence likely achieved polysynaptically - see e.g. Anderson et al., 2016). Specifically, these analyses favored models wherein right MFG modulated all three MTL

regions - parahippocampus, hippocampus, and amygdala - beyond simpler models that excluded the amygdala as a target. This superiority arose even though model selection in DCM penalizes for unnecessary model complexity (Stephan et al., 2007). Though the functional significance of amygdala is not limited to processing aversive information (e.g. Bonnet et al., 2015; Weymar and Schwabe, 2015), our findings nonetheless indicate that reduced amygdala activity when people suppressed unpleasant images in this and prior studies (Depue et al., 2007) likely reflected inhibition in service of affect regulation rather than simply being a downstream effect of memory control.

**Reactive control and supramodal inhibition.** The current data support a tight connection between intrusions and inhibitory modulation of the MTL. Indeed, intrusions triggered stronger down-regulation in all three regions that we studied, replicating findings in the hippocampus when people suppress neutral word pairs (Levy and Anderson, 2012) or scenes (Benoit et al., 2015), and extending this pattern to the amygdala when people suppress negative scenes. Moreover, a frontally mediated inhibitory signal appeared to drive these effects, as reflected in greater negative coupling between the right MFG and MTL sub-regions during intrusions. In fact, Non-Intrusions generated no reliable evidence of negative coupling. These findings suggest that inhibitory control mechanisms were engaged reactively to disrupt momentary awareness of unwanted memories.

Our evidence for parallel modulation of the hippocampus and amygdala echoes findings observed when people suppress visual object memories. Gagnepain et al., (2014) found that suppressing such memories modulated not only the hippocampus, but also fusiform cortex regions involved in object perception. Similarly, we found down-regulation in the parahippocampus and the amygdala, previously associated with scene and emotion reactivation, respectively (Staresina et al., 2012; Smith et al., 2006). Together, these findings suggest that suppression modulates cortical sites specific to the content being controlled, and that modulation is triggered by intrusion-related reinstatement of activity that is then rapidly truncated (Hellerstedt et al., 2016). Based on this “reinstatement principle” (Gagnepain et al., 2014; Hu et al., 2017), suppression should down-regulate amygdala activity and play an affect-regulation function whenever the memories being suppressed elicit emotional responses reinstated by reminders, prompting the targeting of that structure for control.

Recent evidence suggests the existence of a supramodal inhibitory control process spanning motor, memory, and emotion stopping, within the right anterior MFG (Depue et al., 2015). The present findings establish novel support for this claim. Depue et al.’s (2015) claims were based on a conjunction analysis performed on the inhibition contrast for three stopping tasks spanning these domains. The co-localized right MFG activations suggest that components of the stopping mechanisms engaged in these tasks are domain general. In this study, our PLS analysis identified individual voxels that jointly predicted both affect suppression scores and intrusion regulation. Strikingly, the anterior middle frontal gyrus region identified here closely resembles the supramodal region identified by Depue et al.’s conjunction analysis. Our study also adds evidence that this MFG region is effectively connected to both the hippocampus and the amygdala when people suppress unpleasant scenes, consistent with a single mechanism supporting memory control and affect regulation.

The foregoing shared mechanism hypothesis is consistent with the observation that the likely pathways supporting the regulation of the amygdala and the hippocampus transit through common relays including the ventral portion of the medial prefrontal cortex (vmPFC) and the dorsal ACC (Anderson et al., 2016; Ray and Zald, 2012; Thayer et al., 2009; 2012). Nevertheless, aspects of the network implementing each ability surely differ. The MFG has been proposed to reactively interrupt ongoing retrieval in the hippocampus through the thalamus (Anderson et al., 2016) or to regulate an anticipated emotional state (Etkin et al.,



2015) via the ventral portion of the lateral PFC (i.e. IFG), which provides direct input to the amygdala (Ray and Zald, 2012). Thus, even if common control processes in the right MFG mediate affect regulation and mnemonic control, the targets of modulation and the pathways achieving it must diverge.

**Implications.** One finding with important implications is the observation of substantial individual differences in the affective consequences of retrieval suppression. Although suppression reduced amygdala activation when unpleasant scenes intruded, our participants, as a whole, did not show diminished negative affect for suppressed items. However, whereas participants with effective memory control (i.e., fewer intrusions) showed affect suppression, others with poor memory control (more intrusions) exhibited the opposite pattern: increased negative affect for suppressed compared to baseline scenes. These differences suggest that the overall null effect mixes people for whom suppression did and did not work well. Corroborating this interpretation, we found that the better that individuals were at controlling intrusions, the stronger was their engagement of inhibitory control, as reflected by (a) greater right MFG engagement during intrusions, and (b) larger intrusion-related down-regulations in the hippocampus and amygdala. Critically, effective connectivity analyses established that controlling intrusions triggered greater negative coupling between the MFG and these MTL regions, confirming the involvement of inhibitory modulation. This observation suggests that individual differences in inhibitory control may play a critical role in whether suppressing intrusions is an effective coping strategy. Consistent with this, increasing memory inhibition ability (indexed by suppression-induced forgetting or, electrophysiologically, by the suppression-related N2 component) predicts reduced distress from intrusive memories during the week following viewing of a traumatic film (Streb et al., 2016).

We expected that suppressing negative scenes would be harder because we assumed that emotion renders unpleasant memories especially intrusive. Our findings did not confirm this expectation. Neither overall intrusion frequency nor the decline in intrusions over suppressions varied with valence; if anything, participants reported fewer intrusions for negative scenes (Figure 2). Moreover, inhibitory control regions were not reliably more engaged for negative scenes, nor were MTL regions more suppressed. It is unclear why valence did not affect intrusion control. One likely possibility is that our training carefully matched learning for neutral and negative associations, possibly eliminating differences in encoding strength that usually favor emotional materials. Indeed, failure to match encoding quality may partially account for why valence effects on retrieval suppression have proven inconsistent, with some studies showing greater suppression for negative content (Depue et al., 2006; Lambert et al., 2010), and others finding less (Butler and James, 2010; Chen et al., 2012; Norby et al., 2010) or no difference (van Schie et al., 2013; Murray et al., 2011; Murray et al., 2015).

By clarifying how retrieval suppression contributes to affect regulation, the present findings may offer insights into the mechanisms underlying intrusive symptoms in psychiatric disorders. We found that successfully suppressing negative intrusions also regulates emotional responses to the intruding memory; this likely reflects the suppression of affective traces. If so, diminished suppression ability should render people vulnerable not only to recurring intrusions but also to persisting emotional responses that cascade into further distress. Indeed, difficulty controlling upsetting images frequently initiates psychopathological symptoms in disorders including post-traumatic stress and obsessive-compulsive disorders (Brewin et al., 2010). Moreover, compromised suppression-induced forgetting occurs in people with PTSD (Catarino et al., 2014), high ruminators (e.g., Fawcett, et al., 2015), high trait anxiety (Marzi et al., 2014), and low resistance to stress (Lemoult et al., 2010). Some evidence even indicates that attempts to suppress thoughts in such populations can be counter-productive in some



1048 circumstances (see, e.g., Davies and Clark, 1998; Shiperd and Beck, 2005). If deficient  
1049 retrieval suppression contributes to such adverse outcomes, this raises the hope that  
1050 interventions focused on training the mechanisms identified here could, in principle, reduce  
1051 intrusions while dampening negative affect linked to suppressed images.  
1052

1053 **References**

1054

1055 Abdi H, Williams LJ, Beaton D, et al. (2012) Analysis of regional cerebral blood flow data to  
1056 discriminate among Alzheimer's disease frontotemporal dementia and elderly  
1057 controls: multi-block barycentric discriminant (MUBADA) methodology. *J Alzheimer*  
1058 *Dis* 31:S189-S201.

1059 Anderson MC, Bunce J, Barbas H (2016) Prefrontal–hippocampal pathways underlying  
1060 inhibitory control over memory. *Neurobiol Learn Mem* 134:145-161.

1061 Anderson MC, Green C (2001) Suppressing unwanted memories by executive control. *Nature*  
1062 410:366–369.

1063 Anderson MC, Hanslmayr S (2014) Neural mechanisms of motivated forgetting. *Trends Cog*  
1064 *Science* 18:279–292.

1065 Anderson MC, Ochsner K, Kuhl K, Cooper J, Robertson E, Gabrieli SW, Gabrieli JDE (2004)  
1066 Neural systems underlying the suppression of unwanted memories. *Science* 303:232-  
1067 235.

1068 Benoit R, Anderson MC (2012) Opposing mechanisms support the voluntary forgetting of  
1069 unwanted memories. *Neuron* 76: 450-460.

1070 Benoit RG, Davies DJ, Anderson MC (2016) Reducing future fears by suppressing the brain  
1071 mechanisms underlying episodic simulation. *PNAS* 113: E8492–E8501.

1072 Benoit R, Hulbert J, Huddleston E, Anderson MC (2015) Adaptive top-down suppression of  
1073 hippocampal activity and the purging of intrusive memories from consciousness. *J*  
1074 *Cogn Neurosci* 27:96-111.

1075 Bonnet L, Comte A, Tatu L, Millot J, Moulin T, Medeiros de Bustos E (2015) The role of the  
1076 amygdala in the perception of positive emotions: an “intensity detector”. *Front Behav*  
1077 *Neurosci* 9.

1078 Brewin C, Gregory J, Lipton M, Burgess N (2010) Intrusive images in psychological  
1079 disorders: characteristics neural mechanisms and treatment implications. *Psychol*  
1080 *Review* 117:210-232.

1081 Butler AJ, James KH (2010) The neural correlates of attempting to suppress negative versus  
1082 neutral memories. *Cogn. Affect. Behav. Neurosci.* 10:182–194.

1083 Catarino A, Küpper CS, Werner-Seidler A, Dalgleish T, Anderson MC (2015) Failing to  
1084 forget: inhibitory-control deficits compromise memory suppression in posttraumatic  
1085 stress disorder. *Psychol Sci* 26:604-16.

1086 Chen C, Liu C, Huang R, Cheng D, Wu H, Xu P, Mai X, Luo YJ (2012) Suppression of  
1087 aversive memories associates with changes in early and late stages of neurocognitive  
1088 processing. *Neuropsychologia* 50:2839–2848.

1089 Comte M, Schön D, Coull J, Reynaud E, Khalfa S, Belzeaux R, Ibrahim E, Guedj E, Blin O,  
1090 Weinberger D, Fakra E (2014) Dissociating bottom-up and top-Down mechanisms in  
1091 the cortico-limbic system during emotion processing. *Cereb Cortex* 26:144-155.

1092 Davies MI, Clark DM (1998) Thought suppression produces a rebound effect with analogue  
1093 post-traumatic intrusions. *Behav Res Ther* 36:571-582.

- 1094 Depue B, Banich M, Curran T (2006) Suppression of emotional and nonemotional content in  
1095 memory: effects of repetition on cognitive control. *Psychol Sci* 17:441-447
- 1096 Depue B, Burgess G, Willcutt E, Ruzic L, Banich M (2010) Inhibitory control of memory  
1097 retrieval and motor processing associated with the right lateral prefrontal cortex:  
1098 evidence from deficits in individuals with ADHD. *Neuropsychologia* 48:3909-3917.
- 1099 Depue B, Curran T, Banich M (2007) Prefrontal regions orchestrate suppression of emotional  
1100 memories via a two-phase process. *Science* 317:215-219.
- 1101 Depue B, Orr J, Smolker H, Naaz F, Banich M (2015) The organization of right prefrontal  
1102 networks reveals common mechanisms of inhibitory regulation across cognitive  
1103 emotional and motor processes. *Cereb Cortex* 26:1634-1646.
- 1104 Eippert F, Veit R, Weiskopf N, Erb M, Birbaumer N, Anders S (2007) Regulation of  
1105 emotional responses elicited by threat-related stimuli. *Human Brain Mapping* 28:409-  
1106 423.
- 1107 Etkin A, Büchel C, Gross J (2015) The neural bases of emotion regulation. *Nat Rev Neurosci*  
1108 16:693-700.
- 1109 Fawcett J, Benoit R, Gagnepain P, Salman A, Bartholdy S, Bradley C, Chan DKY, Roche A,  
1110 Brewin CR, Anderson MC (2015) The origins of repetitive thought in rumination:  
1111 Separating cognitive style from deficits in inhibitory control over memory. *J Behav*  
1112 *Ther Exp Psychiatry* 47: 1-8.
- 1113 Franko E, Insausti AM, Artacho-Perula E, Insausti R, Chavoix C (2012) Identification of the  
1114 human medial temporal lobe regions on magnetic resonance images. *Hum Brain Mapp*  
1115 35:248-256
- 1116 Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19:1273-  
1117 1302.
- 1118 Gagnepain P, Henson RN, Anderson MC (2014) Suppressing unwanted memories reduces  
1119 their unconscious influence via targeted cortical inhibition. *PNAS* 111:E1310-9.
- 1120 Hayes J, Morey R, Petty C, Seth S, Smoski M, McCarthy G, LaBar K (2010) Staying cool  
1121 when things get hot: emotion regulation modulates neural mechanisms of memory  
1122 encoding. *Front Hum Neurosci* 4:230.
- 1123 Hellerstedt R, Johansson M, Anderson MC (2016) Tracking the intrusion of unwanted  
1124 memories into awareness with event-related potentials. *Neuropsychologia* 89:510-523.
- 1125 Hu X, Bergström ZM, Gagnepain P, Anderson MC (2017) Suppressing Unwanted Memories  
1126 Reduces Their Unintended Influences. *Curr Dir Psychol Sci* 26:197-206.
- 1127 Keightley ML, Chiew KS, Anderson JAE, Grady CL (2011). Neural correlates of recognition  
1128 memory for emotional faces and scenes. *Soc Cogn Affect Neurosci* 6:24-37.
- 1129 Kim S, Hamann S (2007) Neural correlates of positive and negative emotion regulation. *J*  
1130 *Cogn Neurosci* 19:776-798.
- 1131 Kohn N, Eickhoff S, Scheller M, Laird A, Fox P, Habel U (2014) Neural network of cognitive  
1132 emotion regulation — An ALE meta-analysis and MACM analysis. *Neuroimage*  
1133 87:345-355.
- 1134 Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial Least Squares (PLS) methods  
1135 for neuroimaging: a tutorial and review. *Neuroimage* 56:455-75.

- 1136 Lambert AJ, Good KS, Kirk IJ (2010) Testing the repression hypothesis: effects of emotional  
1137 valence on memory suppression in the think-no think task. *Conscious Cogn* 19:281–  
1138 293.
- 1139 Lang PJ, Bradley MM, Cuthbert BN (2008) International affective picture system (IAPS):  
1140 affective ratings of pictures and instruction manual. Technical Report A-8. University  
1141 of Florida Gainesville FL.
- 1142 Lemoult J, Hertel PT, Joormann J (2010). Training the forgetting of negative words: the role  
1143 of direct suppression and the relation to stress reactivity. *Appl Cognit Psychol* 24:365–  
1144 375.
- 1145 Levy B, Anderson MC (2012) Purging of memories from conscious awareness tracked in the  
1146 human brain. *J Neurosci* 32:16785-16794.
- 1147 Leys C, Ley C, Klein O, Bernard P, Licata L (2013) Detecting outliers: do not use standard  
1148 deviation around the mean use absolute deviation around the median. *J Exp Soc*  
1149 *Psychol* 49:764–766.
- 1150 McIntosh A, Lobaugh N (2004) Partial least squares analysis of neuroimaging data:  
1151 applications and advances. *Neuroimage* 23:S250-S263.
- 1152 Marzi T, Regina A, Righi S (2014) Emotions shape memory suppression in trait anxiety.  
1153 *Front Psychol* 4:1001.
- 1154 Moritz S, Hörmann C, Schröder J, Berger T, Jacob G, Meyer B, Holmes E, Späth C,  
1155 Hautzinger M, Lutz W, Rose M, Klein J (2013) Beyond words: sensory properties of  
1156 depressive thoughts. *Cogn Emot* 28:1047-1056.
- 1157 Murray BD, Anderson MC, Kensinger EA (2015) Older adults can suppress unwanted  
1158 memories when given an appropriate strategy. *Psychol Aging* 30:9-25.
- 1159 Murray BD, Muscatell KA, Kensinger EA (2011) Effects of emotion and age on performance  
1160 during a think/no-think memory task. *Psychol Aging* 26:940–955.
- 1161 Nørby S, Lange M, Larsen A (2010) Forgetting to forget: On the duration of voluntary  
1162 suppression of neutral and emotional memories. *Acta Psychol* 133:73-80.
- 1163 Ochsner K, Ray R, Cooper J, Robertson E, Chopra S, Gabrieli J, Gross J (2004) For better or  
1164 for worse: neural systems supporting the cognitive down- and up-regulation of  
1165 negative emotion. *Neuroimage* 23:483-499.
- 1166 Penny W, Stephan K, Daunizeau J, Rosa M, Friston K, Schofield T, Leff A (2010) Comparing  
1167 families of dynamic causal models. *PLoS Comput Biol* 6:e1000709.
- 1168 Poppenk J, Evensmoen H, Moscovitch M, Nadel L (2013) Long-axis specialization of the  
1169 human hippocampus. *Trends Cogn Sci* 17:230-240.
- 1170 Pruessner JC, Li LM, Serles W, Pruessner M, Collins DL, Kabani N, Lupien S, Evans AC  
1171 (2000) Volumetry of hippocampus and amygdala with high- resolution MRI and three-  
1172 dimensional analysis software: minimizing the discrepancies between laboratories.  
1173 *Cereb Cortex* 10:433– 442.
- 1174 Pruessner JC, Kohler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL,  
1175 Evans AC (2002) Volumetry of temporopolar, perirhinal, entorhinal and  
1176 parahippocampal cortex from high-resolution MR images: considering the variability  
1177 of the collateral sulcus. *Cereb Cortex* 12:1342–1353.

- 1178 Rachman S (2007) Unwanted intrusive images in obsessive compulsive disorders. *J Behav*  
 1179 *Ther Exp Psychiatry* 38:402–410.
- 1180 Radaelli D, Sferrazza Papa G, Vai B, Poletti S, Smeraldi E, Colombo C, Benedetti F (2015)  
 1181 Fronto-limbic disconnection in bipolar disorder. *Eur Psychiatry* 30:82-88.
- 1182 Ray R, Zald D (2012) Anatomical insights into the interaction of emotion and cognition in the  
 1183 prefrontal cortex. *Neurosci Biobehav Rev* 36:479-501.
- 1184 Rousseeuw PJ, and Van Drissen K (1999) A fast algorithm for the minimum covariance  
 1185 determinant estimator. *Technometrics* 41 :212–223.
- 1186 Shipperd JC, Beck JG (2005) The role of thought suppression in posttraumatic stress disorder.  
 1187 *Behavior Therapy* 36:277–287.
- 1188 Smith A, Stephan K, Rugg M, Dolan R (2006) Task and content modulate amygdala-  
 1189 hippocampal connectivity in emotional retrieval. *Neuron* 49:631-638.
- 1190 Staresina B, Henson R, Kriegeskorte N, Alink A (2012) Episodic reinstatement in the medial  
 1191 temporal lobe. *J Neurosci* 32:18150-18156.
- 1192 Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ (2017) Dynamic  
 1193 causal models of neural system dynamics: current state and future extensions. *J Biosci*  
 1194 32:129-144.
- 1195 Streb M, Mecklinger A, Anderson MC, Lass-Hennemann J, Michael T (2016) Memory  
 1196 control ability modulates intrusive memories after trauma. *J Affect Disord* 192 :134-  
 1197 142.
- 1198 Thayer J, Åhs F, Fredrikson M, Sollers J, Wager T (2012) A meta-analysis of heart rate  
 1199 variability and neuroimaging studies: Implications for heart rate variability as a marker  
 1200 of stress and health. *Neurosci Biobehav Rev* 36:747-756.
- 1201 Thayer J, Lane R (2009) Claude Bernard and the heart–brain connection: Further elaboration  
 1202 of a model of neurovisceral integration. *Neurosci Biobehav Rev* 33:81-88.
- 1203 Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer  
 1204 B, Joliot M (2002) Automated Anatomical Labeling of activations in SPM using a  
 1205 macroscopic anatomical parcellation of the MNI MRI single-subject brain.  
 1206 *Neuroimage* 15:273-289.
- 1207 Van Essen DC, Dickson J, Harwell J, Hanlon D, Anderson CH, Drury HA (2001) An  
 1208 integrated software system for surface-based analyses of cerebral cortex. *Journal Am*  
 1209 *Med Inform Assoc* 8:443-459.
- 1210 van Schie K, Geraerts E, Anderson MC (2013) Emotional and non-emotional memories are  
 1211 suppressible under direct suppression instructions. *Cogn Emot* 27:1122–1131.
- 1212 Van Schie K, Anderson, MC (2017). Successfully controlling intrusive memories is  
 1213 harder when control must be sustained, *Memory*,  
 1214 DOI:10.1080/09658211.2017.1282518
- 1215 Weymar M, Schwabe L (2016) Amygdala and Emotion: The Bright Side of It. *Front Neurosci*  
 1216 10.
- 1217



1218 **Captions**

1219 **Figure 1. Experimental phases and Hypothesized Dynamics** - (A) After learning face-  
 1220 scene pairs (Negative or Neutral), participants were scanned during the Think/No-Think  
 1221 (TNT) task. For Think items (bounded by green box), participants recalled the associated  
 1222 scene; for No-Think items (bounded by red box), they tried to stop the memory of the scene  
 1223 from entering awareness. Baseline cues were not presented during this TNT phase. Next,  
 1224 participants performed a speeded associative recognition task followed by a self-assessment-  
 1225 manikin (SAM) valence rating task on all picture categories (Think, No-Think, and Baseline)  
 1226 to evaluate how suppression affected memory and emotional perceptions, respectively. On  
 1227 these final tests, baseline items provided an estimate of memory or affect, given that neither  
 1228 retrieval nor suppression has been performed in the interim. Note that the pictures displayed  
 1229 here are not issued from IAPS database but are free-use pictures taken from the internet for  
 1230 illustrative purpose. (B) The dynamics predicted to bring about parallel inhibition of memory  
 1231 and emotion for intrusive negative scenes. Cue input to the hippocampus is predicted to drive  
 1232 pattern completion, followed by recurrent reactivation of scene and emotional features in  
 1233 parahippocampal cortex and amygdala respectively; intrusion-related reactivation in these  
 1234 regions is predicted to trigger parallel inhibition by the right middle frontal gyrus (MFG).  
 1235 Note that we do not propose that MFG directly inhibits these structures given the weak  
 1236 anatomical projections between MFG and amygdala (Anderson et al., 2015). However, the  
 1237 MFG is proposed to modulate these regions polysynaptically, via pathways yet to be fully  
 1238 understood.

1239  
 1240 **Figure 2. Behavioral and Neural indices of mnemonic and affective regulation** - (A)  
 1241 Intrusion proportions (i.e., the proportion of trials in which the associated memory entered  
 1242 into awareness on No-Think trials as measured by our trial-by-trial intrusion report measure -  
 1243 see Procedure) over the 5 scanning blocks of the TNT phase. Shaded error bands represent  
 1244 within-participant standard deviations. (B) The relationship between intrusion proportion and  
 1245 affect suppression score (No-Think – Baseline) for Negative scenes (left panel). Participants  
 1246 who were better at controlling intrusions of unpleasant scenes also showed reduced negative  
 1247 feelings toward them afterwards (right panel). Stars indicate significant differences ( $P < .05$ ).  
 1248 (C) Brain areas more engaged by retrieval suppression than by retrieval (No-Think > Think;  
 1249 hot colors) and vice versa (No-Think < Think; cold colors), thresholded at the uncorrected  
 1250 level of  $P < .001$  for visualization purposes. Pink spheres represent right MFG ROI foci across  
 1251 participants used in subsequent activation, correlation, and DCM analyses (see Method).  
 1252 These ROIs were derived from individual local maxima centered around the supramodal  
 1253 control system described in Depue et al. (2015) and are projected onto a common standard  
 1254 space for visualization purposes. Statistical parametric maps were rendered on the top of the  
 1255 PALS human surface using Caret software (Van Essen et al., 2001; RRID:SCR\_006260).  
 1256

1257 **Figure 3. MTL down-regulation** - (A) Suppressing scene memories reduced activity across  
 1258 the whole MTL overall. Additionally, we observed more pronounced down-regulation in  
 1259 these MTL regions during suppression attempts that were accompanied by intrusions. (B)  
 1260 Distribution of MTL ROI foci across participants once projected back to MNI space. Error  
 1261 bars reflect standard error of the mean (SEM).  
 1262

1263 **Figure 4. Brain/Behavior correlations** - Pearson correlation (skipping bivariate outliers)  
 1264 between affect suppression/intrusion proportion and neural marker of memory suppression  
 1265 (Intrusion - Non-Intrusion) for each scene type. Note that for neural markers of memory  
 1266 suppression in the MTL, negative (lower) scores are assumed to indicate more successful

suppression of activity (Intrusion – Non-Intrusion is a down-regulation in most cases, suggesting control); in contrast, neural markers for memory suppression in MFG, are up-regulations of activity (again, Intrusion – Non-Intrusion) and so positive scores are assumed to indicate greater engagement of control. Higher behavioral scores for affect suppression scores indicate greater reduction in negative valence for No-Think compared to Baseline items (i.e., No-think – Baseline); in contrast, higher intrusion scores, indicate worse control over intrusions. Taken together, these considerations indicate that in MTL, positive correlations for intrusions signify that down-regulations predict fewer intrusions, whereas negative correlations for affect suppression signify that down-regulations predict reduced affect. In MFG, the direction of correlations is expected to invert, because greater control is indicated by up-regulation. Error bars reflect 99.3 % bootstrapped Confidence Interval (CI) corrected for multiple comparisons across ROIs. Significant correlations occur when the CI does not encompass zero.

**Figure 5. Relationship between neural markers of inhibitory control and the reduction of mnemonic awareness/affective response** - Outcome of the PLS analysis for both Negative and Neutral scenes (conducted within the retrieval suppression network, see Method) between intrusion-related up-regulation (Intrusion – Non-Intrusion) and behavioral measures (intrusion proportion and affect suppression score). (A) Voxels showing a significant pattern of brain/behavior correlations as revealed by the first (significant) LV were identified using a BSR threshold higher/lower than 1.96/-1.96, respectively (i.e.,  $P < .05$ ). Correlations between participants' brain scores and behavioral measures for the first significant LV are also reported in panel A (error bars indicate bootstrapped 95% CI). Brain scores reflect the contribution of each participant to a given LV. The correlation between brain scores and behaviors thus reveals the meaning of the LV. (B) Scatter plots observed in the right MFG illustrating the relationship captured by PLS analysis between the up-regulation (Intrusion – Non-Intrusion) and behavioral scores for Negative and Neutral scenes. These findings reveal voxels whose up-regulation is associated with reduced intrusion frequency for both Negative and Neutral scenes and also with increased affect suppression score only in the case of Negative scenes (reduced negative affect for suppressed images). BSR maps were rendered on the top of the PALS human surface using Caret software (Van Essen et al., 2001; RRID:SCR\_006260).

**Figure 6. Relationship between MTL down-regulation and reductions of mnemonic awareness / affective response** - Outcome of the behavioral PLS analysis for both Negative and Neutral scenes conducted within the MTL, including bilateral amygdala, hippocampus, and parahippocampal cortex. PLS correlations were tested with the MTL contrast map for the differences between Intrusion and Non-Intrusion conditions on one hand (i.e., down-regulation), and behavioral scores (intrusion proportion and affect suppression, i.e., the reduction of negative feelings for No-Think items relative to Baseline) on the other hand. (A) Voxels significantly associated to the first significant LV and whose down-regulation significantly correlated with intrusion proportion for both Negative and Neutral scenes, as well as with affect suppression exclusively for Negative scenes (error bars indicate bootstrapped 95% CI). Voxels were identified using a BSR higher or lower than 1.96/-1.96, respectively (i.e.,  $P < .05$ ). Correlations between participants' brain scores and behaviors for the first significant LV are also reported in panel A. Clusters of BSR exceeding threshold were rendered onto a three-dimensional reconstruction of a standard MTL template. The 3D representation of the MTL was obtained by transforming MTL binary masks into 3D meshes using "Anatomist/BrainVISA" software (<http://www.brainvisa.info/>; RRID:SCR\_007354).

1317 (B) Scatter plots illustrating the relationship captured by PLS analysis between the down-  
 1318 regulation observed in the MTL cluster (see panel A) and behavioral scores for both Negative  
 1319 and Neutral scenes.

1320  
 1321 **Figure 7. DCM model space** – DCM models were organized into two families. The first (the  
 1322 Modulatory family) divided the model space into five sub-groups differing according to  
 1323 whether the intrinsic connection from the right MFG was modulated or not by No-Think trials  
 1324 (modeled as 3-second short-epochs separately for Intrusion and Non-Intrusion of each  
 1325 emotion type). Subgroups 1-5 of this family (rows 1-5) either included modulation on bottom  
 1326 up (e.g., hippocampus to MFG), top-down, or bidirectional connections; or no modulation,  
 1327 but variable afferent input to MTL regions from a source independent of MFG; or no  
 1328 modulation at all (row 5). The second family (the Regulation family) divided the model space  
 1329 into families according to modulatory targets. Subgroups 1-3 of this family (columns 1-3)  
 1330 include the Emotion Regulation family (left, amygdala modulation only), the Memory  
 1331 Regulation family (middle, modulation of hippocampus, parahippocampal cortex, or both)  
 1332 and the Parallel regulation family (right, modulation of amygdala and other memory-related  
 1333 regions). After estimating all 35 models for each participant, we performed the group BMS as  
 1334 implemented in SPM12 (version DCM12 revision 4750; RRID:SCR\_007037). This produces  
 1335 the exceedance probability (i.e., the extent to which each model is more likely than any other  
 1336 considered model) and expected posterior probability (i.e., the probability of a model  
 1337 generating the observed data). Note that by positing connectivity relationships between MFG  
 1338 and these MTL structures in our DCM models, we do not presuppose direct anatomical  
 1339 connections (an assumption that DCM's analytical method does not require); rather, we are  
 1340 modeling the data to evaluate the existence of a (potentially) polysynaptic and directional  
 1341 causal influence of each region on the activity of others to which it is connected.

1342  
 1343  
 1344 **Table 1. Behavioral data for valence rating and recognition tasks.** Adjusted valence rating  
 1345 and recognition reaction times (RT) and accuracy for Negative and Neutral scenes (SD are in  
 1346 parentheses).

1347  
 1348 **Table 2. Regions showing a difference in activity between Think and No-Think trials.**  
 1349 Peak coordinates of the regions showing greater activity for No-Think relative to Think items  
 1350 (and *vice versa*) across Negative and Neutral scenes at  $P_{FWE} < 0.05$  (whole-brain).

1351  
 1352 **Table 3. Control network regions showing a significant pattern of brain/behavior**  
 1353 **correlations as revealed by the first latent variable of the PLS analysis.** Peak coordinates  
 1354 of the control regions significantly loading on the first LV. Rightmost columns indicate the  
 1355 correlation between cluster up-regulation (Intrusion *versus* Non-Intrusion) and intrusion  
 1356 proportion/affect suppression score.

1357  
 1358 **Table 4. MTL regions showing a significant pattern of brain/behavior correlations as**  
 1359 **revealed by first latent variable of PLS analysis.** Peak coordinates of the MTL regions  
 1360 significantly loading on the first LV. Rightmost columns indicate the correlation between  
 1361 cluster down-regulation (Intrusion *versus* Non-Intrusion) and intrusion proportion/affect  
 1362 suppression score.

1363  
 1364 **Table 5. DCM intrinsic and modulatory parameters.** Mean DCM.A (intrinsic), and  
 1365 effective connectivity (DCM.B + DCM.A), between MFG and MTL regions for Intrusions (I)  
 1366 and Non-Intrusions (NI) by scene valence (SDs in parentheses). Significant coupling

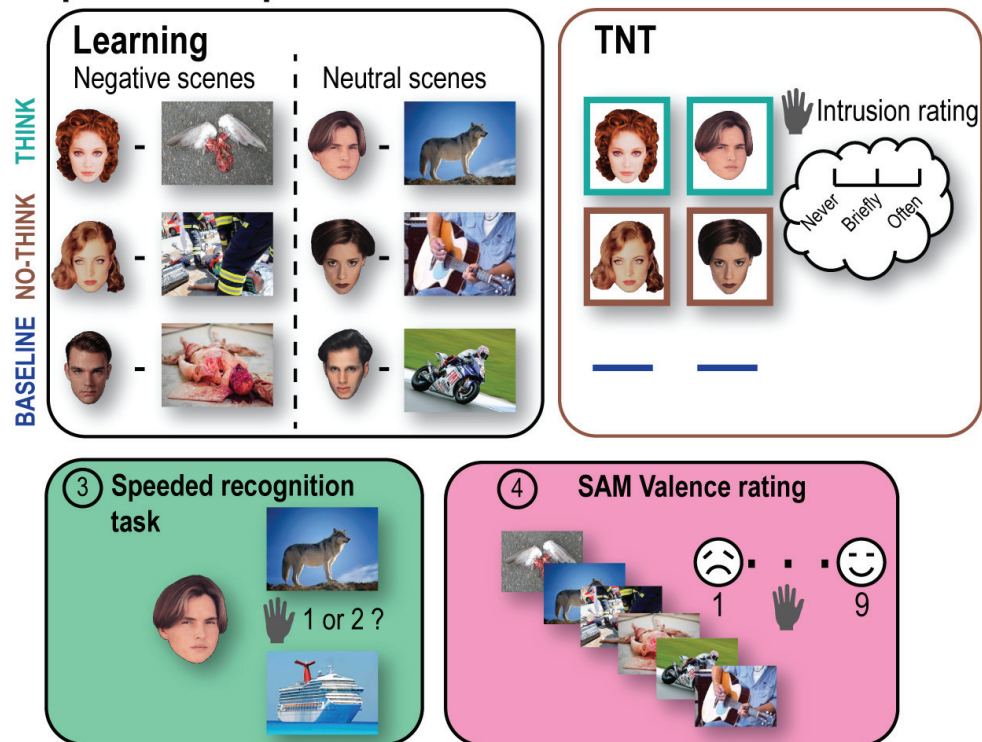
1367 parameters are indicated in bold with an asterisk (99.6% bootstrapped CI corrected for  
1368 multiple comparisons using Bonferroni correction).

1369

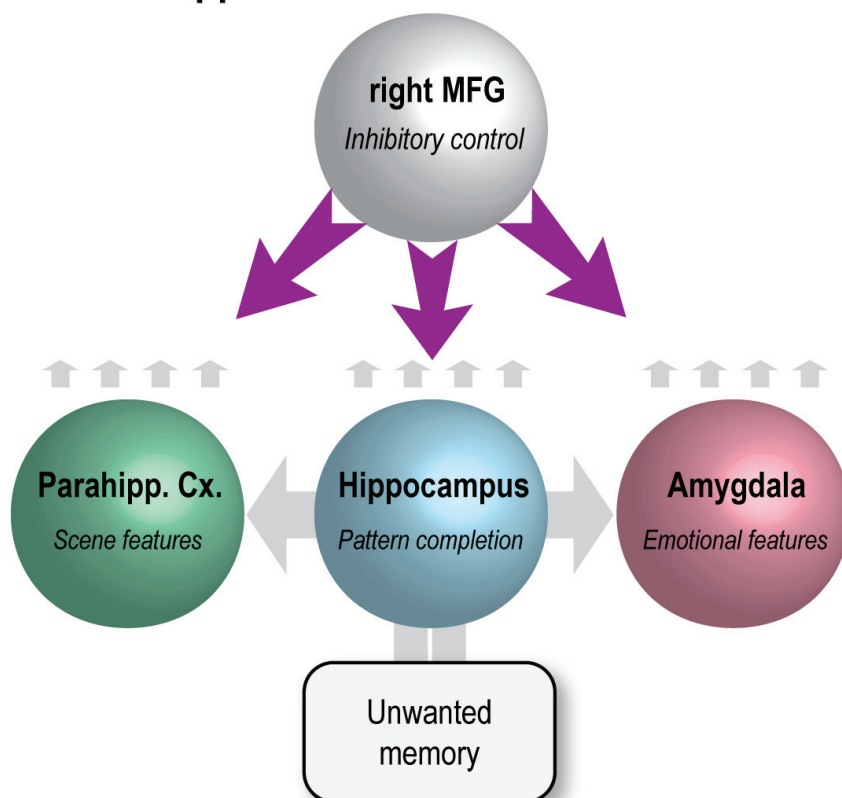
1370

1371

## A Experimental phases

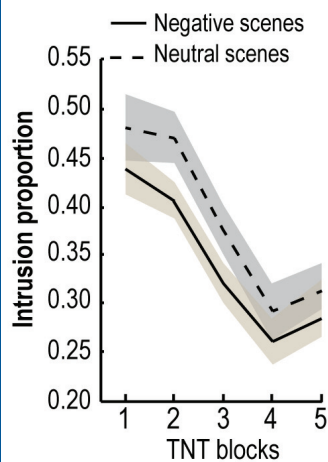


## B Parallel regulation of memory and emotion during retrieval suppression

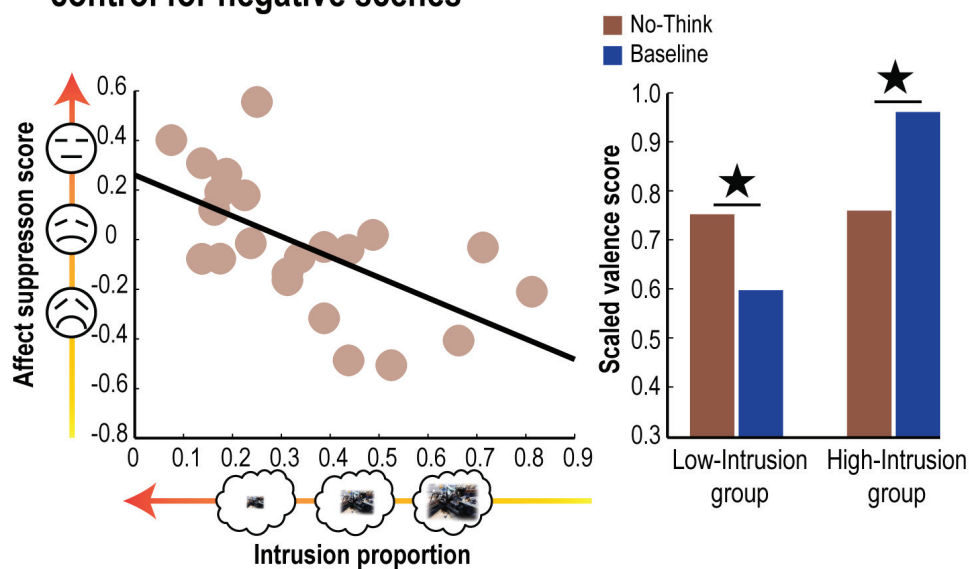




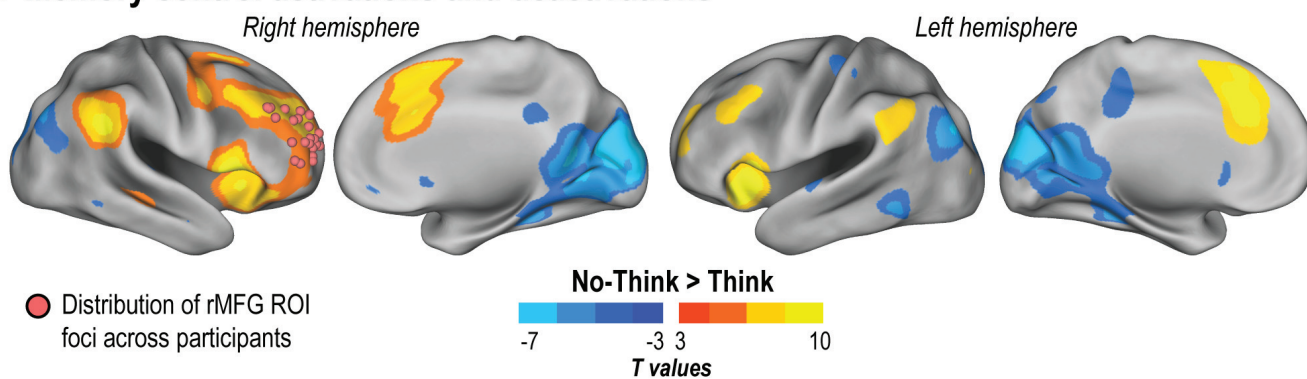
## A Intrusion proportion across TNT blocks



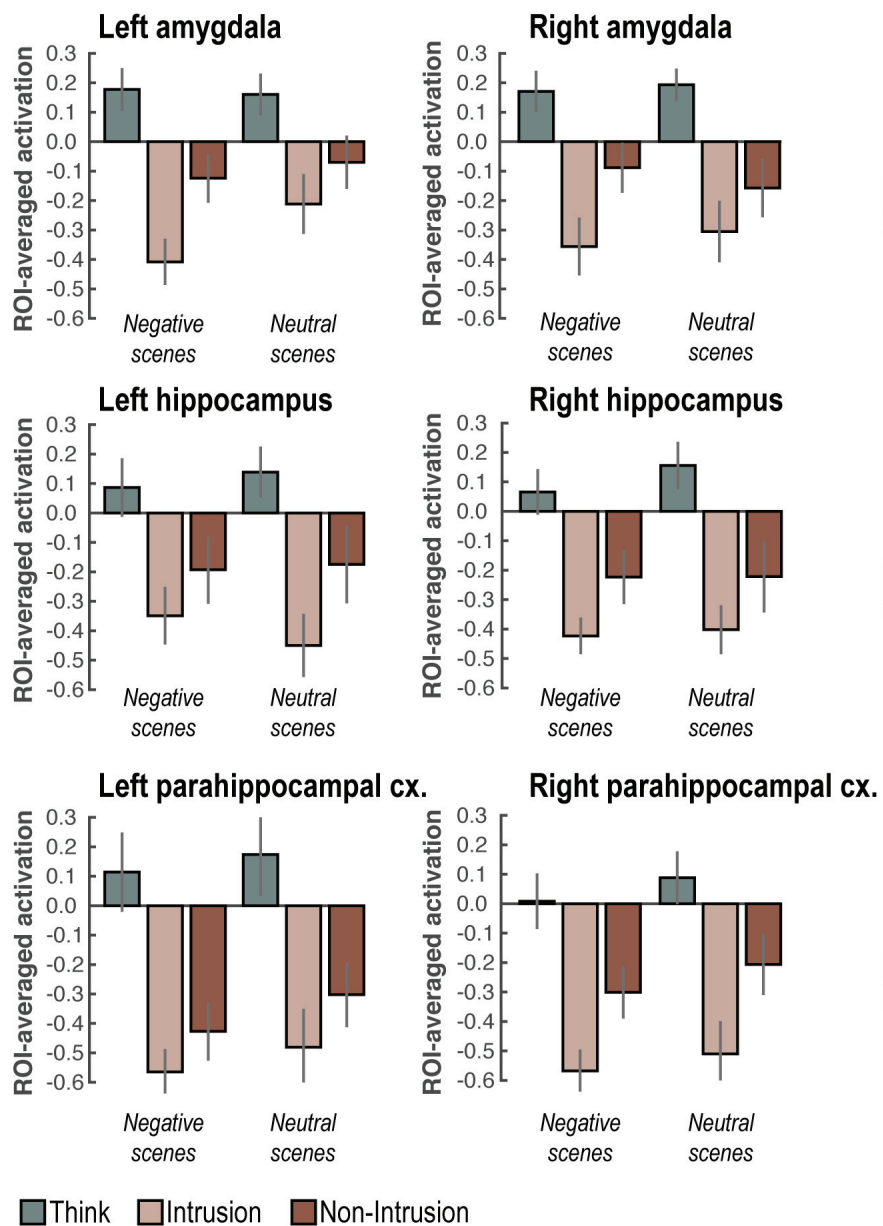
## B Relationship between affect suppression and intrusion control for negative scenes



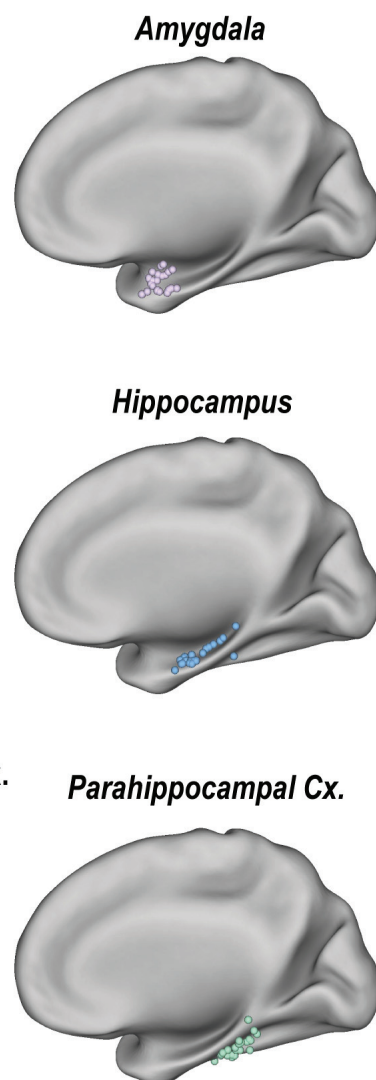
## C Memory control activations and deactivations

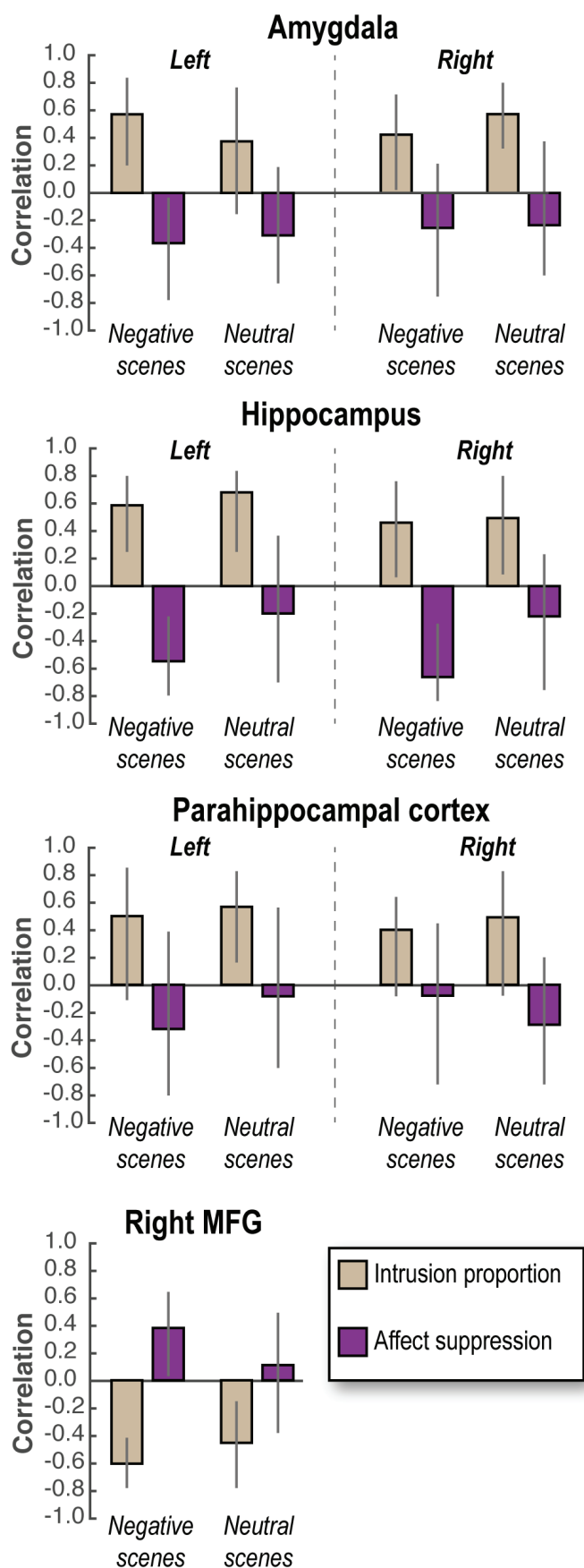


## A Region of Interest (ROI) analyses

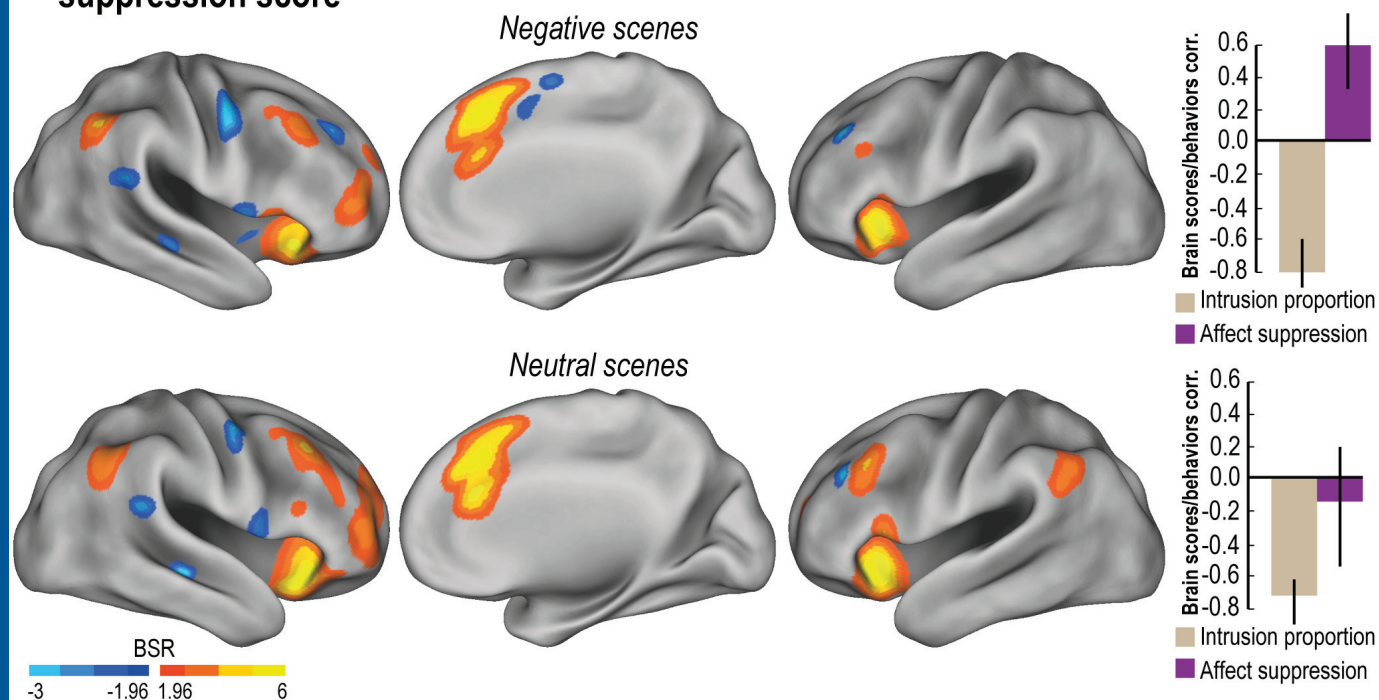


## B Distribution of MTL ROI foci across participants

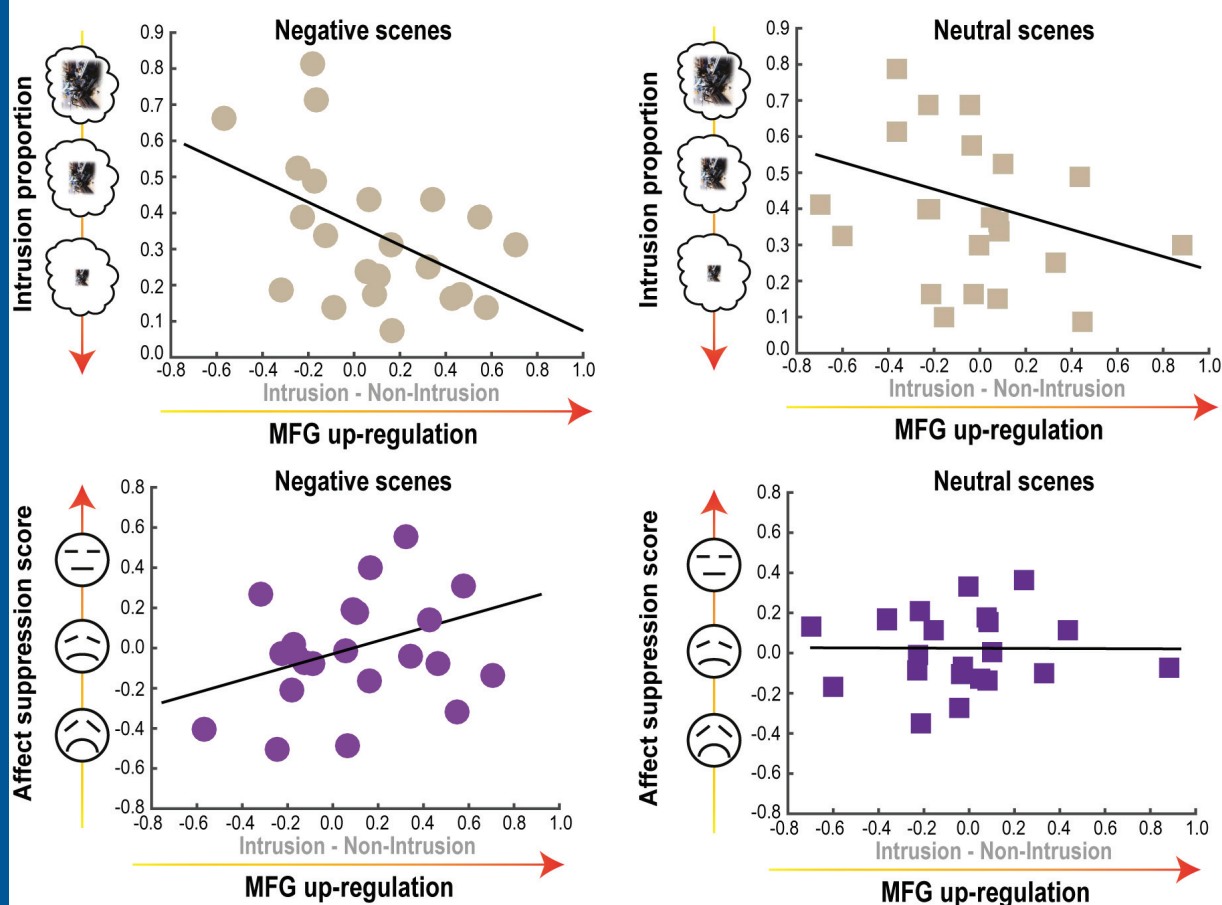




# A Relationship between fronto-parietal up-regulation and intrusion proportion/affect suppression score

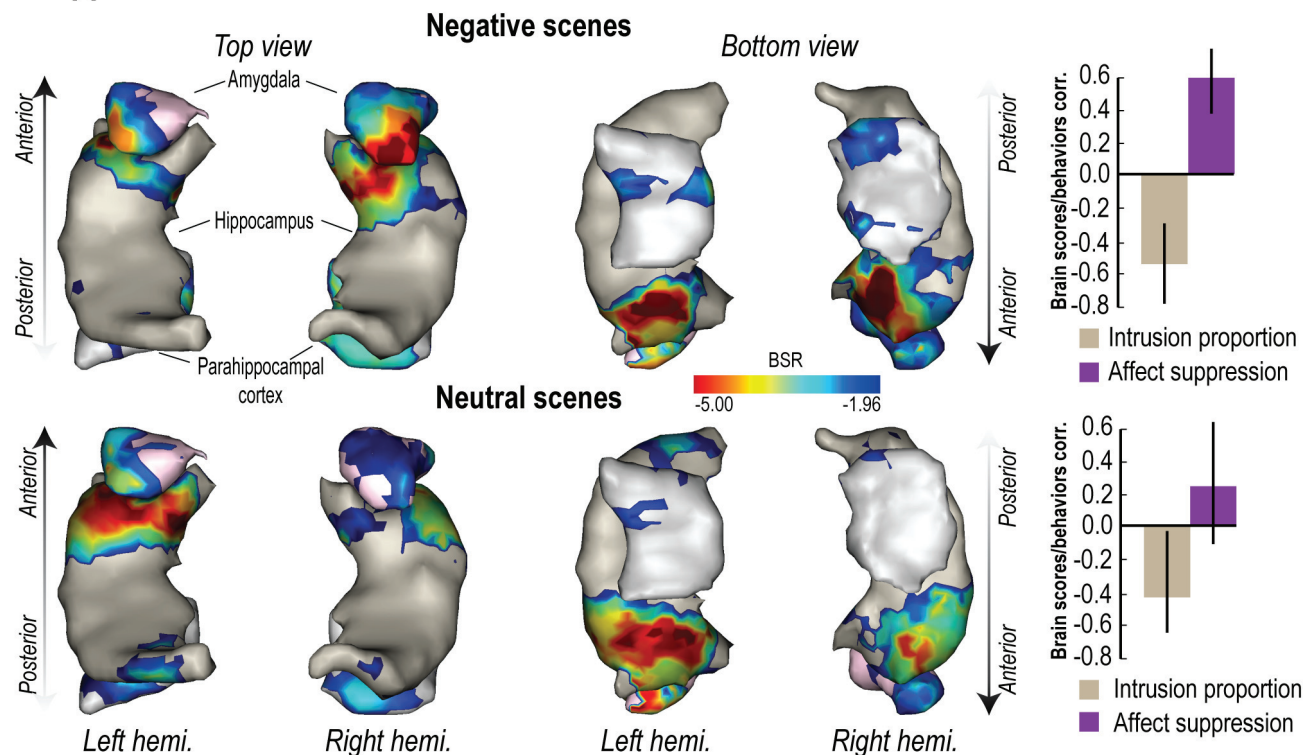


## B Illustration of the relationship between MFG up-regulation and behaviors

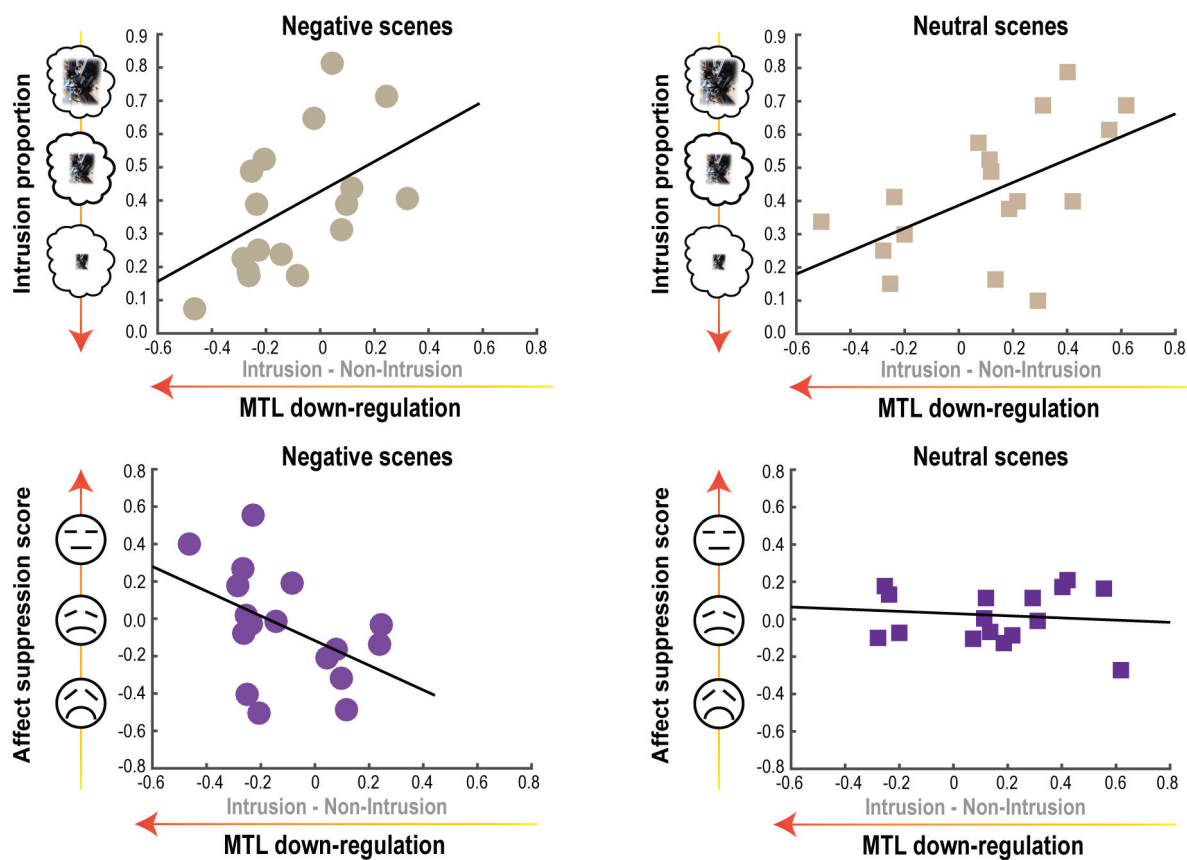




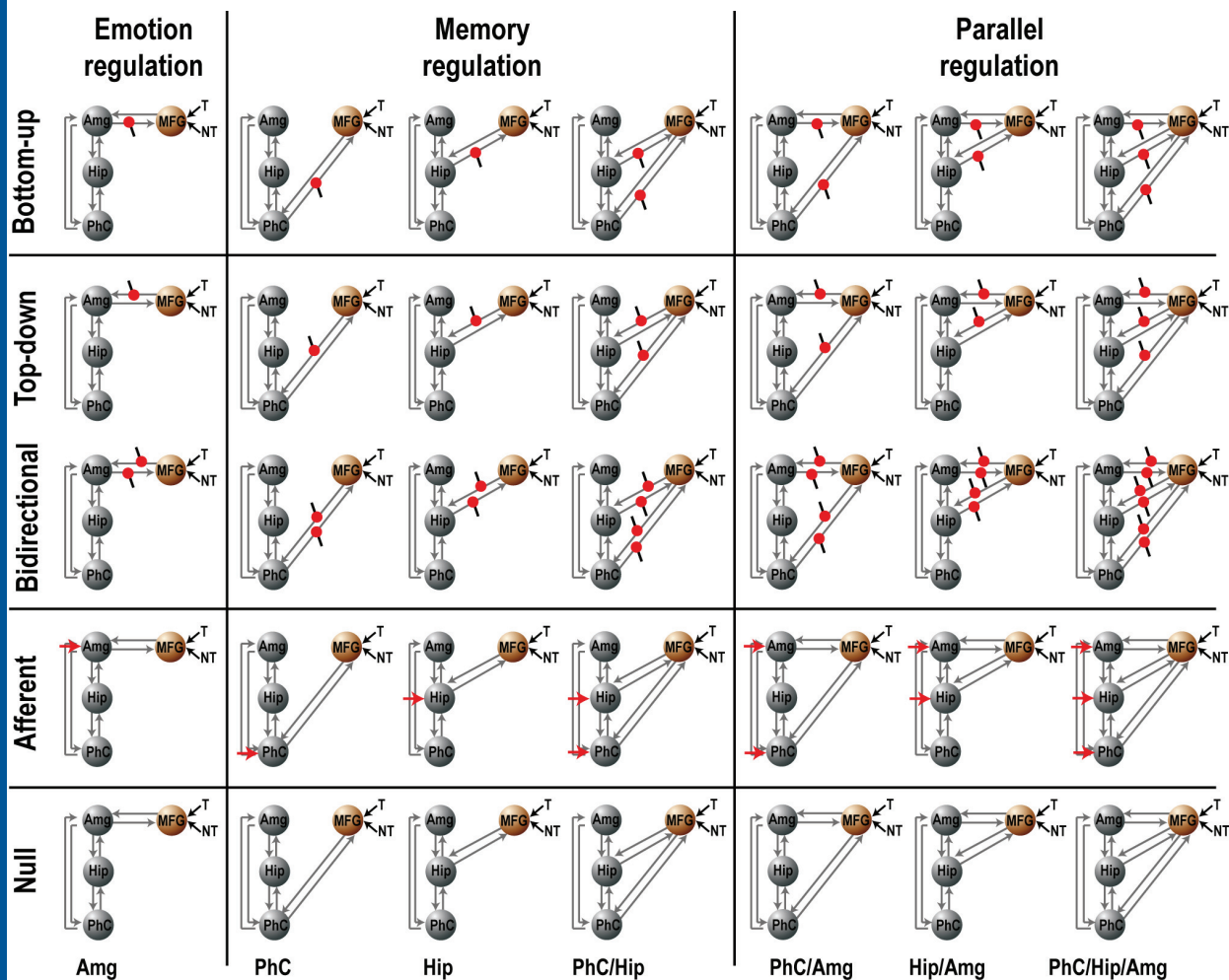
## A Relationship between MTL down-regulation and intrusion proportion/affect suppression score



## B Illustration of the relationship between MTL down-regulation and behaviors







← : model driving inputs

→ : afferent driving inputs during intrusion and non-intrusion of negative and neutral scenes

• : top-down and bottom-up modulatory inputs during intrusion and non-intrusion of negative and neutral scenes

↔ : intrinsic connections

**Table 1. Behavioral data for valence rating and recognition tasks.**

<b>Emotion</b>	<b>Valence rating</b>			<b>Recognition RT (ms)</b>			<b>Recognition Accuracy</b>		
	<b>Think</b>	<b>No-Think</b>	<b>Baseline</b>	<b>Think</b>	<b>No-Think</b>	<b>Baseline</b>	<b>Think</b>	<b>No-Think</b>	<b>Baseline</b>
<i>Negative scenes</i>	0.86 (0.35)	0.76 (0.32)	0.78 (0.37)	1444 (338)	1573 (495)	1442 (438)	0.97 (0.10)	0.97 (0.05)	0.98 (0.07)
<i>Neutral scenes</i>	1.14 (0.15)	1.16 (0.17)	1.14 (0.14)	1320 (313)	1356 (295)	1344 (336)	0.97 (0.05)	0.96 (0.06)	0.97 (0.06)

Table 2. Regions showing a difference in activity between Think and No-Think trials.

Anatomical description	Number of voxels	MNI coordinates (mm)			Z	P <sub>FWE</sub>
		x	y	z		
No-Think > Think						
Right inferior frontal gyrus	1859	48	20	-2	Inf	< .001
Right superior frontal gyrus	2011	14	14	58	7.48	< .001
<i>dACC</i>		4	24	32	6.88	< .001
Right inferior parietal cortex	662	60	-46	32	7.28	< .001
Right middle frontal gyrus	1591	42	24	36	6.34	< .001
<i>Right middle frontal gyrus (anterior)</i>		30	48	16	6.15	< .001
<i>Right middle frontal gyrus (anterior)</i>		28	48	32	5.86	< .001
<i>Right middle frontal gyrus (posterior)</i>		44	14	41	5.74	< .001
<i>Right precentral gyrus</i>		50	2	40	5.29	< .005
Left insula	753	-46	16	0	7.04	< .001
<i>Left inferior frontal gyrus</i>		-56	20	2	5.96	< .001
Left middle frontal gyrus	23	-28	52	14	4.93	< .01
Right superior temporal gyrus	21	54	-28	10	4.74	< .05
Left inferior parietal cortex	6	-52	-58	44	4.70	< .05
Right striatum	18	16	2	8	4.64	< .05
Think > No-Think						
Left Cuneus	4387	0	-84	24	7.47	< .001
Left middle occipital gyrus	209	-38	-78	32	6.31	< .001
Left post-central gyrus	19	-40	-24	62	4.70	< .05
Right fusiform gyrus	28	32	-36	-16	5.04	< .005
Left fusiform gyrus	19	-30	-38	-14	4.65	< .05

**Table 3. Control network regions showing a significant pattern of brain/behavior correlations as revealed by the first latent variable of the PLS analysis.**

Anatomical description	N voxels	MNI coordinates (mm)			BSR	P value	R-Intrusion proportion	R-Affect suppression
		x	y	z				
Negative scenes								
Left superior frontal medial gyrus	1865	2	24	44	9.42	0.0000000	-0.7822	0.5294
Right insula/inferior frontal gyrus	1072	28	24	0	6.95	0.0000000	-0.6010	0.5700
Left Insula/inferior frontal gyrus	520	-32	18	-2	5.40	0.0000000	-0.5193	0.4027
Right inferior parietal cortex	182	52	-54	46	4.82	0.0000007	-0.6991	0.3424
Right middle frontal gyrus (ant.)	215	38	50	8	4.67	0.0000015	-0.5491	0.4638
Right middle frontal gyrus (post.)	143	40	24	42	4.08	0.0000228	-0.5648	0.2996
Left middle frontal gyrus	10	-34	50	4	2.99	0.0013989	-0.4767	0.2461
Right superior frontal gyrus	29	22	56	24	2.40	0.0082399	-0.3670	0.4384
Right middle frontal gyrus	67	32	36	46	-4.27	0.0000099	0.4337	-0.4310
Right precentral gyrus	108	50	-4	50	-4.16	0.0000157	0.6450	-0.5005
Right rolandic operculum	52	56	6	2	-3.59	0.0001665	0.4331	-0.4357
Right supplemental motor area	27	8	-2	58	-3.12	0.0009183	0.4253	-0.4943
Right superior temporal gyrus	67	54	-40	22	-3.06	0.0010909	0.3099	-0.4957
Right middle temporal gyrus	23	48	-22	-8	-2.52	0.0059107	0.2889	-0.4060
Neutral scenes								
Right dACC	2465	6	28	34	10.2	0.0000000	-0.6823	-0.2088
Right insula/inferior frontal gyrus	1099	34	20	-14	10.1	0.0000000	-0.6789	-0.1806
Right middle frontal gyrus (post.)	444	38	24	36	7.39	0.0000000	-0.5318	-0.1780
Left insula/inferior frontal gyrus	906	-30	18	-16	5.80	0.0000000	-0.5960	-0.0471
Right middle frontal gyrus (ant.)	611	40	54	4	5.77	0.0000000	-0.5453	-0.2580
Right angular gyrus	335	48	-58	40	4.59	0.0000023	-0.6438	-0.0113
Left middle frontal gyrus	129	-38	26	34	3.95	0.0000393	-0.4905	-0.2614
Right superior frontal gyrus	17	18	56	-8	3.38	0.0003675	-0.4326	0.0491
Right superior frontal gyrus	17	18	6	56	3.11	0.0009341	-0.2907	-0.1668
Left supramarginal gyrus	131	-60	-54	30	3.03	0.0012034	-0.4391	0.1107
Right striatum	23	16	12	0	2.86	0.0020956	-0.3100	-0.1602
Left superior frontal gyrus	11	-26	54	16	2.57	0.0050181	-0.3600	0.0185
Right middle temporal gyrus	13	60	-32	-8	2.47	0.0066930	-0.4098	0.0604
Right precentral gyrus	44	46	-6	48	-2.85	0.0021916	0.5611	-0.2745
Right superior temporal gyrus	20	50	-26	-4	-2.47	0.0067875	0.4536	-0.2466
Right supramarginal gyrus	13	56	-36	30	-2.39	0.0083475	0.4167	-0.1571
Right superior temporal gyrus	17	56	-44	18	-2.31	0.0105426	0.4501	-0.2108

**Table 4. MTL regions showing a significant pattern of brain/behavior correlations as revealed by first latent variable of PLS analysis.**

Anatomical description	N voxels	MNI coordinates (mm)			BSR	P value	R-Intrusion proportion	R-Affect suppression
		x	y	z				
Negative scenes								
Right hippocampus/amygdala	580	24	-12	-18	-7.04	0.0000000	0.5268	-0.5867
Left hippocampus/amygdala	336	-30	-8	-16	-6.18	0.0000000	0.5113	-0.5489
Right parahippocampal cortex	202	20	-36	-14	-4.25	0.0000109	0.4385	-0.6719
Left parahippocampal cortex	178	-20	-32	-14	-4.19	0.0000141	0.6096	-0.4909
Neutral scenes								
Left amygdala/hippocampus	487	-30	0	-22	-6.89	0.0000000	0.3916	-0.2846
Right hippocampus/amygdala	499	34	-18	-8	-6.63	0.0000000	0.3566	-0.2604
Right parahippocampal cortex	97	28	-38	-12	-4.67	0.0000015	0.3007	-0.3681
Left hippocampus/parahipp. cortex	173	-20	-34	0	-4.16	0.0000162	0.4173	-0.2385



**Table 5. DCM intrinsic and modulatory parameters.**

	DCM.A	DCM.A + DCM.B			
	Intrinsic	Negative		Neutral	
		I	NI	I	NI
MFG → Phc	-0.22 (0.52)	<b>-0.55*</b> (0.69)	-0.36 (0.81)	<b>-0.49*</b> (0.76)	-0.14 (0.54)
MFG → Hip	-0.16 (0.35)	<b>-0.32*</b> (0.54)	-0.25 (0.62)	<b>-0.32*</b> (0.52)	-0.19 (0.48)
MFG → Amg	-0.09 (0.29)	<b>-0.47*</b> (0.63)	-0.13 (0.60)	-0.18 (0.97)	-0.39 (0.61)