Purging of Memories from Conscious Awareness Tracked in the Human Brain

Benjamin Levy
University of San Francisco, bjlevy3@usfca.edu

Michael C. Anderson

Follow this and additional works at: http://repository.usfca.edu/psyc

Part of the Neurology Commons, and the Psychology Commons

Recommended Citation

This Article is brought to you for free and open access by the College of Arts and Sciences at USF Scholarship Repository. It has been accepted for inclusion in Psychology by an authorized administrator of USF Scholarship Repository. For more information, please contact zjlu@usfca.edu.
Behavioral/Systems/Cognitive

Purging of Memories from Conscious Awareness Tracked in the Human Brain

Benjamin J. Levy1 and Michael C. Anderson2

1Department of Psychology, University of San Francisco, San Francisco, CA 94117-1080 and 2MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom

Understanding the neural basis of conscious experience and its regulation are fundamental goals of science. While recent research has made substantial progress in identifying the neural correlates of conscious experiences, it remains unclear how individuals exert control over the contents of awareness. In particular, can a memory that has entered the aware state be purged from consciousness if it is not currently desired? Here we tracked the correlates of consciousness in humans using functional magnetic resonance imaging and demonstrated the involvement of a downregulation mechanism that purges contents from conscious awareness. When individuals tried to prevent the retrieval of a memory in response to reminders, hippocampal activity was reduced, as previously established. Crucially, using trial-by-trial reports of phenomenal awareness, we found that this reduction of hippocampal activation was specifically associated with moments when a memory involuntarily intruded into conscious awareness and needed to be purged. This downregulation of activity during memory intrusions appears to disrupt momentary awareness of unwanted contents and, importantly, predicts impaired recall of the memory on later tests. These results tie the voluntary control of phenomenal awareness to observable changes in neural activity linked to awareness, and so provide a neurobiological model for guiding inquiry into the physical foundations of control over consciousness.

Introduction

Neurobiological research on consciousness has revealed neural activity that tracks the entrance of perceptions and memories into awareness (Crick and Koch, 1995; Kreiman et al., 2002; Rees et al., 2002). For example, both single unit electrophysiology (Kreiman et al., 2000; Gelbard-Sagiv et al., 2008) and functional neuroimaging (Eldridge et al., 2000; Eichenbaum et al., 2007; Montaldi and Mayes, 2010) have identified neural activity in the human hippocampus that tracks the presence of memories in conscious awareness, indicating that activity in this structure contributes to memories achieving the aware state (Clark and Squire, 1998). It is less clear, however, how control over the contents of awareness is achieved. Here we monitored neural markers of awareness to examine whether people could make contents leave consciousness voluntarily, and, if so, how this purging is accomplished in the brain. We hypothesized that the controlled suppression of neural activity supporting awareness may be a key mechanism supporting the purging of unwanted contents.

To test this hypothesis, we focused on the hippocampus to determine whether neural activity contributing to mnemonic awareness can be suppressed when people purge its contents. We examined how people control mnemonic awareness using the Think/No-Think (TNT) paradigm developed to study retrieval stopping (Anderson and Green, 2001). Prior work with this paradigm demonstrates that attempting to stop retrieval in response to a cue makes it harder to recall the associated memory on later tests (Anderson and Green, 2001; Anderson and Huddy, 2012), a finding hypothesized to reflect the weakening of suppressed traces through inhibitory control (Anderson and Green, 2001). We propose that this inhibitory control process, triggered by the detection of unwanted traces in awareness, suppresses hippocampal activity, providing a physical basis by which people purge an experience from consciousness.

To examine the purging of content from phenomenal awareness, we isolated moments when a memory entered a person’s awareness and this awareness needed to be suppressed, and further linked these perceptions of memory to functional magnetic resonance imaging (fMRI) measures of hippocampal activation. For this purpose we used the trial-by-trial introspection method developed in research on attention (Sergent and Dehaene, 2004; Sergent et al., 2005; Corallo et al., 2008). This method uses reports of private, first-person experience collected immediately after a cognitive operation to analyze behavioral and neuroimaging data according to phenomenological state. Following this method, our participants classified their experience after each trial in the TNT task according to whether the presented cue led its associated memory to enter consciousness (Fig. 1A). We made four predictions. First, memories often would intrude into awareness involuntarily, but attempts to stop retrieval would decrease intrusion frequency. Second, how quickly intrusions declined would predict later memory suppression, reflecting a relationship between the processes that exclude a memory from consciousness and forgetting.
Third, consciously recollecting associates on Think trials would increase hippocampal activation, confirming its validity as a marker of mnemonic awareness. Finally, and importantly, we predicted that during No-Think trials, because we asked participants to exclude the associate from awareness, unintended awareness of a memory would trigger inhibitory control to counteract hippocampal signals that contribute to the recollective state. If so, tracking hippocampal activation during these experiences may allow us to witness the neural basis of people’s control over awareness in action.

Materials and Methods

Participants

Eighteen people (10 female, age range: 18–33) participated. All participants had spoken English as a primary language since early childhood and had no history of attention deficit disorder or attention deficit hyperactivity disorder. Two additional participants participated but were excluded from the study, one due to scanner error and one for failure to follow the experimental instructions. An additional 96 participants were enrolled in a behavioral pilot study.

Materials

Stimuli consisted of 57 word pairs, adapted from Anderson et al. (2004). Thirty-six experimental word pairs were divided into three sets, which rotated across participants through the conditions (Think, No-Think, and Baseline). The remaining pairs were fillers, which were used as buffers at the beginning and end of lists to avoid primacy and recency effects, and for practice trials.

Procedure

The procedure had three phases: study, TNT, and test. All phases were performed inside the scanner, but fMRI data were only acquired during the TNT phase.

Study phase. Participants studied cue-associate word pairs for 5 s each and then practiced retrieving the responses. On each trial, the cue appeared and participants had 5 s to recall the associate and say it out loud. After every trial, the correct associate word appeared as feedback for 2 s. An experimenter used the keyboard to indicate if each response was correct. If the correct associate was not provided, that cue was re-presented after the other pairs had been tested until each associate was recalled correctly once. Participants were then tested once more on each pair, without feedback, to confirm pairs had been learned. Performance on this test was high (96%).

TNT. On each trial, participants received the cue from a pair (e.g., Needle) and were asked either to recall and think about the associate (e.g., Doctor; Think condition) or to prevent the associate from entering consciousness (No-Think condition). Each cue word appeared for 3 s. After every trial, participants received up to 1.5 s to report their experience of whether the associate entered awareness during that trial by pressing the appropriate button on an fMRI-compatible box with their right hand. Participants were urged to make these reports honestly and accurately, and it was emphasized that inaccurate classifications would be harmful to the experiment. Participants were further asked to make their classifications quickly and intuitively. Participants performed two practice blocks (25 trials each): first, without making awareness reports, and then with the reports. Both practice blocks were interrupted midway through by the experimenter, who administered a structured interview, which included corrective feedback, to ensure that participants performed the task as requested. The actual TNT phase consisted of six blocks of 54 trials (324 trials total), with each lasting ~6 min. Each block included two repetitions of each experimental word pair (12 Think cues and 12 No-
Behavioral measures

The final test data were analyzed using a mixed-design ANOVA with the type of final test (same probe or independent probe) and type of item (Think, Baseline, or No-Think) manipulated within subjects and the counterbalancing of items through each condition (three levels) and the order of presentation for the two final tests (two levels) manipulated between subjects. The behavioral pilot study included an additional between-subjects manipulation (presence or absence of intrusion ratings). Of critical importance to our hypotheses were planned comparisons that contrast recall of No-Think items and Baseline items, which allowed us to assess whether baseline forgetting was observed.

The other primary dependent measure was how frequently subjects reported thinking about No-Think response words either “briefly” or “often” for each of the 12 repetitions of the cue word. A mixed-design ANOVA was performed on the intrusion frequency data with repetition included as a within-subject manipulation (12 levels) and word counterbalancing (3 levels) included as a between-subject manipulation.

To assess individual differences, we calculated two behavioral measures for each participant. The suppression score was calculated by subtracting recall of No-Think items from Baseline items and then averaging this measure across both of the final tests to provide a measure of how much worse participants’ memories were for associates they had avoided. Note that this subtraction treats forgetting as a positive value, so subjects who showed more forgetting of the avoided memories had higher suppression scores. The intrusion slope score was calculated by taking the frequency of intrusions. Both measures were proportionalized on initial intrusion frequency to account for pretrial variability. These values were then entered into a region by hemisphere by condition ANOVA to test for significance across participants. Given our a priori interest in hippocampal activation, however, we begin by reporting planned comparisons within bilateral hippocampus (we had no a priori theories about hemispheric asymmetries) and then proceed to discuss the pattern in other MTL regions. These same values were also used to study the correlation with behavioral measures of forgetting.

Results

Behavioral results

Memory performance on the final test

Consistent with prior findings (Anderson and Green, 2001; Anderson and Huddleston, 2012), suppressing retrieval of No-Think items harmed people’s later memory for those items on the final test, relative to Baseline pairs that were also studied but that did not appear during the TNT phase (Fig. 1B). In particular, all three groups showed significantly more forgetting of No-Think items than Baseline items (fMRI subjects: \( F_{(1,12)} = 16.5, p < 0.005 \); control subjects with intrusion ratings: \( F_{(1,84)} = 8.6, p < 0.005 \); control subjects without intrusion ratings: \( F_{(1,84)} = 9.6, p < 0.005 \)) and this did not interact with test type (fMRI subjects: \( F_{(1,12)} = 2.4, p = 0.15 \); control subjects with intrusion ratings: \( F_{(1,84)} = 1.4, p = 0.24 \); control subjects without intrusion ratings: \( F_{(1,84)} = 0.1, p = 0.031 \)). Furthermore, in the control study this forgetting effect did not interact with whether or not participants had to classify in-trial intrusive reports to recollect memories formed at study, exceeding what would be expected based on the passage of time, and forgetting did not.
interact with whether participants reported subjective ratings of their mnemonic experience. These findings confirm that efforts to control conscious awareness during retrieval suppression impair the later ability to voluntarily recreate consciousness of memories on a delayed test (Anderson and Green, 2001).

In contrast to the No-Think items, engaging in retrieval did not, in general, affect retention of Think items ($M = 86\%$) compared to performance in the Baseline condition ($M = 87\%$), either in our whole sample of 114 participants ($F_{(1,96)} = 1.8$, $p = 0.18$) or when the analysis was restricted to participants who performed subjective ratings ($F_{(1,96)} = 1.5$, $p = 0.22$), nor was there an interaction between subjects who reported ratings and those who did not ($F_{(1,96)} < 1$). As might be expected, No-Think items were also impaired relative to Think items, across all 114 participants ($F_{(1,96)} = 13.4$, $p < 0.001$) as well as within the more restricted sample that reported subjective ratings, ($F_{(1,96)} = 7.5$, $p < 0.01$), again, this did not interact with whether or not subjects reported subjective ratings ($F_{(1,96)} < 1$). It is perhaps surprising that retrieval did not facilitate performance, though facilitation effects are quite small in this paradigm in general, particularly when performance is near to ceiling, and no feedback is given during the TNT phase (Levy and Anderson, 2008).

**Reports of conscious awareness during retrieval suppression**

During the TNT phase, participants reported that associates were successfully brought into awareness on nearly all of the Think trials ($M = 97.1\%$, Fig. 1C and Table 1). On No-Think trials, however, participants reported that associates entered awareness far less frequently ($M = 37.5\%$), showing that participants were largely successful at regulating awareness according to task goals ($t_{(17)} = 11.98$, $p < 0.001$). Nevertheless, No-Think associates frequently intruded into conscious awareness. Of the intrusions that occurred, 93% were classified as "brief" (Table 1), reflecting either the truncation of mnemonic awareness by control or a failure to sustain attention toward that memory. Importantly, we predicted that memories would frequently intrude into awareness involuntarily initially, but that with repeated attempts to stop retrieval, intrusion frequency would decline. Indeed, intrusions declined appreciably from the first suppression attempt to the twelfth (fMRI subjects: $F_{(1,15)} = 18.5$, $p < 0.001$; control subjects: $F_{(1,45)} = 48.6$, $p < 0.001$) indicating that participants gained increasing control over the entrance of memories into conscious awareness (Fig. 1D). This improved regulation of awareness may reflect a mixture of increasing success at preempting retrieval and accumulating inhibition of suppressed traces that renders them less intrusive over trials.

**Relationship between the control of awareness and later forgetting**

If successfully excluding a memory from conscious awareness and later forgetting represent outcomes of a common inhibitory control process, one should find that variations in regulating awareness predict the amount of forgetting. Specifically, we predicted that the rate at which intrusions declined over repetitions would predict later memory suppression, reflecting a relationship between the control mechanisms that exclude a memory from consciousness and later forgetting. Consistent with this hypothesis, the participants who improved the most at reducing intrusions over repetitions also experienced the most suppression-induced forgetting on the final test ($r = 0.45$, $p < 0.05$; Fig. 1E). There was also a correlation between the overall frequency of intrusions and forgetting ($r = 0.62$, $p < 0.01$), again suggesting that the subjects who were best able to overcome intrusions were the ones who were most likely to forget. This hypothesized control process, triggered to suppress momentary conscious awareness of an involuntary recollection, may achieve this regulatory outcome in part by disrupting the hippocampal representation that contributes to mnemonic awareness, and that also supports the ability to recreate awareness voluntarily on later tests. Importantly, these findings indicate that moments of unwanted awareness indeed arise in this procedure, that people actively suppress that awareness as we requested, and that we can isolate those moments on a trial-by-trial basis to track neural activity relating to the purging of contents from consciousness.

**fMRI results**

**Brain regions engaged by the attempt to regulate conscious**

Attempts to stop retrieval in the No-Think condition has previously been found to engage a frontotoparietal network and to reduce activation in the hippocampus (Anderson et al., 2004; Depue et al., 2007). Consistent with these observations, activity during No-Think trials, relative to Think trials, revealed widespread prefrontal engagement of both lateral prefrontal (including dorsolateral and ventrolateral prefrontal cortex and lateral premotor cortex) and medial prefrontal regions (including anterior cingulate cortex, supplementary motor area, and presupplementary motor area), along with bilateral intraparietal sulcus (Fig. 2A; Tables 2, 3). Importantly, this evidence for the engagement of cognitive control during No-Think trials was accompanied by reduced activation in the bilateral hippocampus and parahippocampal cortex, relative to Think trials. Thus, the expected neural systems engaged by retrieval suppression were found, despite the inclusion of trial-by-trial queries of people’s conscious experience during Think and No-Think trials.

The modulation of hippocampal activation observed here tracks the robust differences in phenomenological reports of conscious recollection between the Think and No-Think conditions (Fig. 1C), and it is consistent with our assumption that recollection is linked to hippocampal activity. To further explore this modulation, both hippocampi were anatomically defined in each individual participant based on structural landmarks (Insausti et al., 1998; Pruessner et al., 2000, 2002; Zeineh et al., 2000, 2003). Activity within this anatomical ROI was clearly modulated by the task, as activity was greater during Think trials than during No-Think trials ($F_{(1,127)} = 10.1$, $p < 0.01$; Fig. 2B). Importantly, this modulation reflected both increased activation during Think trials ($F_{(1,127)} = 6.9$, $p < 0.05$) and decreased activation during No-Think trials ($F_{(1,127)} = 4.5$, $p < 0.05$), relative to prettrial baseline activity. Thus, hippocampal activation arising in response to a retrieval cue varied according to participants’ recollection goals. Increased hippocampal activation during Think trials confirms the validity as a marker of mnemonic awareness, consistent with prior electrophysiological and hemodynamic findings (Eldridge et al., 2000; Kreiman et al., 2000; Eichenbaum et al., 2007;...
Gelbard-Sagiv et al., 2008; Montaldi and Mayes, 2010). The reduction in hippocampal activation during No-Think trials supports a role of inhibitory control in limiting hippocampal contributions to conscious awareness. Crucially, however, these findings leave unanswered whether hippocampal activity was modulated proactively in response to TNT cues to preempt retrieval, or instead was suppressed reactively to purge memories that momentarily intruded into awareness. Addressing this question requires examining hippocampal activation in relation to reports of mnemonic awareness.

Figure 2. Imaging results and the relationship between hippocampal activation and below-baseline forgetting. A. Comparison of brain activity during Think and No-Think trials. Warm colors indicate more activation when participants tried to keep the associate out of mind than when they tried to bring the associate to mind, whereas cool colors indicate regions that were less active during attempts to suppress conscious awareness (black arrows highlight reduced activity in the hippocampus). B. Activation in a priori structurally defined hippocampal ROIs for Think trials and No-Think trials. C. Activation in a priori structurally defined hippocampal ROIs for Think trials and both types of No-Think trials: intrusions where the to-be-avoided memory entered awareness and nonintrusions where memory retrieval was successfully stopped. D. The magnitude of signal reduction in the hippocampus during intrusions (the average percentage signal change between 4 and 8 s after stimulus onset, displayed as a positive value) was correlated, across participants, with below-baseline forgetting of No-Think items on the final test. E. This same measure of hippocampal activity during nonintrusions trials was not related to the amount of forgetting later observed. Error bars for all panels represent SEM.
Table 2. Regions more active during No-Think trials than during Think trials

<table>
<thead>
<tr>
<th>Brain region</th>
<th>~BA</th>
<th>Max Z</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Cluster extent, voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right inferior frontal gyrus</td>
<td>44</td>
<td>5.85</td>
<td>52</td>
<td>12</td>
<td>−2</td>
<td>32514</td>
</tr>
<tr>
<td>Left orbitofrontal cortex</td>
<td>11</td>
<td>5.70</td>
<td>−34</td>
<td>20</td>
<td>−10</td>
<td>62</td>
</tr>
<tr>
<td>Pre-SMA</td>
<td>8/9</td>
<td>5.54</td>
<td>−2</td>
<td>32</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>Pre-SMA</td>
<td>8</td>
<td>5.52</td>
<td>−6</td>
<td>16</td>
<td>52</td>
<td>34</td>
</tr>
<tr>
<td>Pre-SMA</td>
<td>6/8</td>
<td>5.48</td>
<td>4</td>
<td>12</td>
<td>66</td>
<td>24</td>
</tr>
<tr>
<td>Pre-SMA/dorsal ACC</td>
<td>6/32</td>
<td>5.42</td>
<td>6</td>
<td>22</td>
<td>40</td>
<td>18</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>4.54</td>
<td>−34</td>
<td>−60</td>
<td>−36</td>
<td>14</td>
<td>4792</td>
</tr>
<tr>
<td>Left lateral occipital cortex</td>
<td>18/19</td>
<td>4.38</td>
<td>−30</td>
<td>−88</td>
<td>14</td>
<td>68</td>
</tr>
<tr>
<td>Left lateral occipital cortex</td>
<td>18/19</td>
<td>4.26</td>
<td>−30</td>
<td>−84</td>
<td>−16</td>
<td>82</td>
</tr>
<tr>
<td>Left occipital fusiform gyrus</td>
<td>37/19</td>
<td>4.07</td>
<td>−30</td>
<td>−84</td>
<td>−10</td>
<td>76</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>3.91</td>
<td>−32</td>
<td>−68</td>
<td>−36</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Left cerebellum</td>
<td>3.79</td>
<td>−48</td>
<td>−60</td>
<td>−42</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>4.85</td>
<td>−18</td>
<td>−72</td>
<td>54</td>
<td>2382</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>4.32</td>
<td>−18</td>
<td>−68</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>4.19</td>
<td>−22</td>
<td>−64</td>
<td>60</td>
<td>20</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>4.09</td>
<td>−38</td>
<td>−58</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>3.90</td>
<td>−20</td>
<td>−64</td>
<td>48</td>
<td>16</td>
</tr>
<tr>
<td>Left intraparietal sulcus</td>
<td>7</td>
<td>3.88</td>
<td>−16</td>
<td>−64</td>
<td>56</td>
<td>16</td>
</tr>
<tr>
<td>Right intraparietal sulcus</td>
<td>7</td>
<td>4.51</td>
<td>26</td>
<td>−66</td>
<td>58</td>
<td>2300</td>
</tr>
<tr>
<td>Right intraparietal sulcus</td>
<td>7</td>
<td>4.23</td>
<td>14</td>
<td>−64</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>Right intraparietal sulcus</td>
<td>7</td>
<td>4.15</td>
<td>24</td>
<td>−62</td>
<td>48</td>
<td>16</td>
</tr>
<tr>
<td>Right intraparietal sulcus</td>
<td>7</td>
<td>4.05</td>
<td>18</td>
<td>−70</td>
<td>58</td>
<td>16</td>
</tr>
<tr>
<td>Right supramarginal gyrus</td>
<td>40</td>
<td>3.99</td>
<td>54</td>
<td>−46</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Right intraparietal sulcus</td>
<td>7</td>
<td>3.93</td>
<td>36</td>
<td>−64</td>
<td>58</td>
<td>16</td>
</tr>
</tbody>
</table>

*ACC, anterior cingulate cortex; BA, Brodmann’s area; DLPFC, dorsolateral prefrontal cortex; SMA, supplementary motor area.

This cluster also encompassed several other regions including bilateral portions of DLPFC, basal ganglia, insula, and lateral frontal cortex, as well as left inferior frontal gyrus.

This cluster also extended down into inferior lateral parietal cortex in the supramarginal gyrus.

This cluster also extended down into inferior lateral parietal cortex in the supramarginal gyrus.

Table 3. Regions more active during Think trials than during No-Think trials

<table>
<thead>
<tr>
<th>Brain region</th>
<th>~BA</th>
<th>Max Z</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Cluster extent, voxels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occipital lobe</td>
<td>17</td>
<td>5.16</td>
<td>20</td>
<td>−96</td>
<td>20</td>
<td>11785</td>
</tr>
<tr>
<td>Occipital lobe</td>
<td>17</td>
<td>4.76</td>
<td>22</td>
<td>−94</td>
<td>16</td>
<td>62</td>
</tr>
<tr>
<td>Dorsal posterior cingulate cortex</td>
<td>31</td>
<td>4.58</td>
<td>−6</td>
<td>−34</td>
<td>48</td>
<td>24</td>
</tr>
<tr>
<td>Retrosplenial cortex</td>
<td>29</td>
<td>4.49</td>
<td>10</td>
<td>−62</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Retrosplenial cortex</td>
<td>29</td>
<td>4.38</td>
<td>14</td>
<td>−62</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Retrosplenial cortex</td>
<td>29</td>
<td>4.36</td>
<td>8</td>
<td>−54</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Left insula/Heschl’s gyrus</td>
<td>13/41/42</td>
<td>4.6</td>
<td>−38</td>
<td>−20</td>
<td>2</td>
<td>2892</td>
</tr>
<tr>
<td>Left central operculum</td>
<td>4.38</td>
<td>−46</td>
<td>−18</td>
<td>18</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Left central operculum</td>
<td>4.36</td>
<td>−60</td>
<td>−8</td>
<td>6</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Left Heschl’s gyrus</td>
<td>41/42</td>
<td>4.31</td>
<td>−42</td>
<td>−22</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Left insula</td>
<td>13</td>
<td>4.31</td>
<td>−36</td>
<td>−16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Left planum polare</td>
<td>4.01</td>
<td>−50</td>
<td>−6</td>
<td>2</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Ventral ACC/medial PFC</td>
<td>32/10</td>
<td>4.43</td>
<td>6</td>
<td>48</td>
<td>−2</td>
<td>1592</td>
</tr>
<tr>
<td>Subcallosal cortex</td>
<td>25</td>
<td>4.27</td>
<td>0</td>
<td>20</td>
<td>−10</td>
<td>16</td>
</tr>
<tr>
<td>Subcallosal cortex</td>
<td>25</td>
<td>3.89</td>
<td>4</td>
<td>26</td>
<td>−16</td>
<td>16</td>
</tr>
<tr>
<td>Frontopolar cortex</td>
<td>10</td>
<td>3.7</td>
<td>14</td>
<td>62</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Subcallosal, Medial PFC</td>
<td>25/32</td>
<td>3.65</td>
<td>0</td>
<td>26</td>
<td>−16</td>
<td>16</td>
</tr>
<tr>
<td>Frontopolar cortex</td>
<td>10</td>
<td>3.6</td>
<td>6</td>
<td>56</td>
<td>−4</td>
<td>16</td>
</tr>
</tbody>
</table>

*ACC, anterior cingulate cortex; BA, Brodmann’s area; PFC, prefrontal cortex.

This cluster also extended forward into the right hippocampus [peak: 32, −26, 14] and parahippocampal cortex [peak: 32, −34, −16].

This cluster also extended medially into the left anterior hippocampus [peak: −4, −10, −18].

Reports of conscious awareness and hippocampal modulation

Because participants’ goal on No-Think trials is to stop a retrieval cue from eliciting awareness of an associated memory, any intrusion of that memory into awareness is counter-intentional, and, we hypothesize that such a violation of task goals would be more likely to trigger inhibitory control to terminate awareness. As such, our most important prediction was that during No-Think trials, unintended awareness of a memory would trigger inhibitory control to counteract hippocampal signals that contribute to the recollective state. To observe this proposed purging of memories from awareness, we isolated trials on which participants perceived intruding memories in consciousness, contrary to task goals, and tracked a neural marker of the regulatory impact of control. Given prior observations of downregulated activity in the hippocampus during retrieval suppression (Anderson et al., 2004; Depue et al., 2007) we thought this was a good candidate mechanism for controlling awareness. This possibility is consistent with analogous downregulations observed in neocortical areas representing distracting perceptual information during working memory tasks, which also have been interpreted as markers for inhibitory control (Gazzaley et al., 2005). If this hypothesis is correct, and if inhibitory control is triggered by neural signals of awareness, then we should find more downregulation of hippocampal activation during No-Think trials accompanied by momentary awareness of an intruding memory than during No-Think trials not accompanied by awareness.

Hippocampal activation for intrusions and nonintrusions strongly confirmed this hypothesis. During intrusions, the hippocampus exhibited robust reductions in activation relative to baseline (F(1,17) = 17.3, p < 0.001; Fig. 2C). In the absence of an intrusion experience, however, the reduction in hippocampal activation did not reliably differ from either baseline (F < 1) or Think trials (F(1,17) = 1.5, p > 0.2). Importantly, there was significantly less hippocampal activation during intrusions than during nonintrusions (F(1,17) = 9.9, p < 0.01; Fig. 2C). These findings isolate the negative BOLD response observed during retrieval suppression to trials that require participants to purge conscious awareness of intruding memories.

Notably, reduced hippocampal activation during intrusions is the inverse of what one expects based on the presumed relationship between hippocampal activation and the degree of recollection. Given that intentional recollection (Think trials) increased hippocampal activation as expected, we suggest that this inversion arises from a brief recollective signal that is then countered by a robust inhibitory control response. This recollective signal during intrusions may be difficult to measure given its brevity, variable timing, and limits on the temporal resolution of fMRI. If hippocampal downregulation is triggered as a response to momentary intrusions, these findings associate people’s goal-directed efforts to purge a memory from conscious awareness with measurable neural impacts of control on hippocampal activity.

Hippocampal modulation during conscious intrusions predicts later forgetting

It was necessary to confirm, however, that the reduced activation evident during intrusions truly reflects the hypothesized control mechanism thought to underlie the purging of momentary conscious awareness. To establish this we needed to show that the reduced hippocampal response during intrusions was not merely the absence of hippocampal activation, but rather a functionally important downregulation. Toward that end, we examined whether reductions in hippocampal activation predicted behavioral memory impairment on the final test, which previously had been established to involve inhibitory control (Anderson and Green, 2001; Anderson et al., 2004; Depue et al., 2007; Anderson and Huddleston, 2012). The hypothesized downregulation of hippocampal activity may be especially disruptive when it occurs during an intrusion, in which the neural representation of an
episodic trace has been briefly reactivated (Nader et al., 2000; Dudai, 2004). Inhibitory control may achieve such disruption (Anderson and Spellman, 1995; Anderson, 2003) and if so, the amount of signal reduction during conscious intrusions should predict memory impairment.

Strikingly, the extent to which hippocampal activation was reduced during intrusions predicted later memory suppression robustly \( (r = 0.70, p < 0.005; \text{Fig. 2D}) \). That is, participants who showed the greatest reduction in hippocampal activity during intrusions were the ones who showed the largest below-baseline memory suppression effects for No-Think items. That hippocampal downregulation predicted memory performance that was worse than baseline implies the existence of a mechanism that disrupts retention. In contrast, when intrusions did not occur, variations in the amount of hippocampal activity bore no reliable relationship to forgetting (Fig. 2E).

These findings strongly support the hypothesis that intrusions trigger active downregulation that disrupted later recall of reactivated traces, consistent with reactivation-induced lability (Nader et al., 2000; Dudai, 2004). Noninfiltration trials did not reliably trigger this downregulation response, possibly because items on those trials did not require control or because awareness was successfully preempted by other control processes that did not involve the modulation of hippocampal activity. Collectively, these findings support the view that reduced hippocampal activation during intrusions reflects active downregulation that purges mnemonic contents from awareness. Prior research suggests that a likely candidate for the source of this downregulation is the dorsolateral prefrontal cortex (DLPFC), as engagement of this region during No-Think trials predicts below-baseline forgetting (Anderson et al., 2004; Depue et al., 2007). Consistent with those findings, we observed two clusters that predicted individual differences in below-baseline forgetting (at a relaxed threshold, \( p < 0.005, 20 \) voxel minimum), both of which fell within DLPFC (peak voxels: \( x = -18, y = 44, z = 44; x = 30, y = 28, z = 48 \)). Intriguingly, activity in the left DLPFC cluster during intrusions negatively correlated with right hippocampal activation during intrusions \( (r = -0.48, p < 0.05) \), but only during a later time window \( (6–10 \) s) than the one used in earlier analyses. This apparent negative coupling between DLPFC and hippocampus suggests that the DLPFC may provide the source of top-down control. No reliable correlations with forgetting or hippocampal activity were observed in this region during nonintrusions. These speculations fit well with evidence, recently discovered with dynamic causal modeling, demonstrating that DLPFC is negatively coupled with hippocampus during retrieval suppression, particularly for good inhibitors (Benoit and Anderson, 2012).

Patterns similar to those observed in the hippocampus were also observed in other MTL regions.

Our primary theoretical hypotheses concerned activity within the hippocampus, but it is possible that other regions within the MTL are similarly modulated by this task. In addition to anatomically defining the hippocampus, we also created anatomically defined ROIs for the perirhinal, entorhinal, and parahippocampal cortices. We then repeated our analyses for the hippocampus in these other regions to investigate whether similar patterns are observed in these other MTL regions as well. First, we investigated whether a similar ordering of the Think, nonintron, and intrusion was observed in these regions (Fig. 3). An ANOVA including all four MTL ROIs revealed a significant main effect of condition \( (F_{(2,34)} = 5.9, p < 0.01) \) and a marginal interaction with hemisphere \( (F_{(6,102)} = 5.06, p = 0.06) \), but condition did not interact with region \( (F < 1) \) and the three-way interaction with region and hemisphere was also not significant \( (F_{(6,102)} = 1.6, p = 0.16) \). This suggests that although we have focused on downregulation of hippocampal activation, the control response appears to extend beyond the hippocampus. The marginal interaction with hemisphere reflects the fact that modulation tended to be more robust in the right hemisphere than the left hemisphere, particularly for the cortical regions (hemisphere did not interact with any of the effects reported earlier within the hippocampus, \( F_{3} < 1.5 \)). We also assessed whether the magnitude of downregulation in these regions during intrusions (but not during noninfiltrations) predicted participant’s suppression scores. We again observed similar patterns in several of these other MTL regions (Fig. 3). While these cortical regions were only investigated in a post hoc fashion, it is interesting to note that the control response appears to extend beyond the hippocampus proper into surrounding cortical structures.

**Discussion**

A central feature of consciousness is the ability to voluntarily direct and control its contents. The present results provide unique evidence that the capacity to purge the contents of awareness is achieved by suppressing neural activity underpinning the aware state—in this case, activation in the hippocampus. Critically, this modulation of hippocampal activation during the purging of awareness predicted later retention of the purged trace. These data provide a striking link between the intentional regulation of phenomenal awareness and the neural events that implement this control.

The linkage between the intentional regulation of awareness and the neural manifestations of control observed here is unusually specific. Simply separating No-Think trials according to participants’ first-person experience of whether they briefly perceived a memory in conscious awareness revealed starkly different levels of activation in the hippocampus, allowing us to witness the regulatory response on neural processes contributing to mnemonic awareness. One interpretation of this specificity is that inhibitory control was triggered by the emergence of unwanted memorial contents in consciousness. Uprogration of inhibitory control in reaction to consciously detected conflict is predicted, for instance, by the global neuronal workspace model of consciousness (Dehaene et al., 1998; Sergent and Dehaene, 2004), which posits a sharp nonlinear transition between unconscious and conscious processing, and, importantly, a dependence of executive control on detection of conscious conflict (Dehaene et al., 2003; Mayr, 2004). A second interpretation is that awareness is sufficient, but not necessary to trigger inhibitory control. Inhibitory control may be upregulated in response to unwanted activation regardless of whether a memory is conscious, and reactivation in the hippocampus may simply correlate highly with awareness. Indeed, even noninfiltration trials showed a modest, though not reliable reduction in activity during retrieval suppression, suggesting that awareness may simply increase the level of control. Whichever interpretation is correct, these findings support the view that when unwanted traces are perceived in awareness, inhibitory control can be deployed to suppress activity that contributes to awareness to end this phenomenological state, establishing a link between inhibitory control and the purging of content from consciousness.

Though our data indicate that inhibition is upregulated to purge memories from awareness, inhibition may not be limited to this reactive role. Inhibition may be engaged proactively, for example, to prevent intrusions. Indeed, the sizeable reduction in
Figure 3. Activation in anatomically defined MTL ROIs and the relationship between this activation and suppression scores. While the primary hypotheses concerned the hippocampus, for completeness we report the data from all four MTL ROIs in the right (A) and left (B) hemisphere. Each quadrant represents an ROI (illustrated in the coronal slices in the center of the figure) and within those quadrants are four panels: activity during Think and No-Think trials, without distinguishing between intrusions and nonintrusions (upper left); activity during Think, nonintrusion, and intrusion trials (upper right); the relationship between behavioral suppression scores and downregulation of this region during nonintrusion trials (lower left); and the relationship between suppression scores and downregulation of this region during intrusion trials (lower right). The value of distinguishing awareness of No-Think items is reflected in the increased control response to intrusions throughout the panels, and in its robust prediction of forgetting. All error bars reflect SEM.
awareness of the memory on the very first retrieval suppression trial ($M = 60\%$) compared to the first Think trial ($M = 97\%$) suggests that participants are very good at proactively controlling awareness. This proactive control may be achieved by downregulating neocortical activity in areas not addressed in the current study. Intrusions, by this view, may simply trigger one specific form of inhibitory control that involves modulating activity within the hippocampus, and may not represent the only form of control involved in stopping retrieval. These conjectures fit with the fact that, while hippocampal reductions during intrusions predict a sizable portion of the variance (nearly $50\%$) in below-baseline forgetting, they do not explain all forgetting, suggesting the mechanisms involved in preempting intrusions may also impair memory without hippocampal modulation. Moreover, inhibition is unlikely to be the only means of controlling awareness. Control over awareness may also be sustained over time, for example, by additional mechanisms engaged in parallel with or after inhibition, that redirect attention to alternative foci, such as to the reminder or to substitute thoughts (Benoit and Anderson, 2012). At a minimum, however, the present findings show that hippocampal activation relating to the aware state can be suppressed to disrupt awareness, and that this suppression impairs retention of the suppressed trace.

Perhaps the most compelling aspect of the hippocampal responses observed here is that variations in the degree of down-regulation during intrusions predicted future forgetting of the purged trace. The greater the reduction in hippocampal activation during an intrusion, the worse later memory performance was for No-Think items, compared with performance on Baseline items. This variation in hippocampal downregulation underscores how reduced hippocampal activation is not an obligatory sequela of an intrusion; rather, it reflects the outcome of a control response, varying widely across individuals, which holds the potential to make memory for suppressed items worse than it would ordinarily be. Critically, this relationship of hippocampal modulation with later retention was absent for nonintrusions, linking suppression-induced forgetting specifically to the purging of a reactivated trace from awareness. These observations suggest that reductions in hippocampal activity during No-Think trials reflect active control and not a mere failure to engage the hippocampus. If reduced hippocampal activity during intrusions simply reflected a failure to engage hippocampally mediated recollection, there would be no reason to expect a worsening of memory compared with baseline; rather, memories would merely be deprived of further benefit that would have arisen from their reactivation. Second, a failure to engage the hippocampus during retrieval suppression also has difficulty explaining why hippocampal activation during intrusions was not more similar to that observed during Think trials, owing to increased engagement of recollection on those suppression trials. That intrusions were associated with further reductions in activation compared with nonintrusions suggests an active response counteracting recollection. Thus intrusion-specific reductions in activation are uniquely predictive of forgetting, consistent with active downregulation in service of purging the contents of mnemonic awareness.

The strong and specific relationship between our measure of memory inhibition and hippocampal downregulation during intrusions echoes work showing that reactivating a memory makes it especially vulnerable (Nader et al., 2000; Dudai, 2004). Disruption might, in the present design, occur via the suppression of activity supporting awareness, a possibility consistent with other forms of reactivation-dependent inhibition (Anderson and Spellman, 1995; Anderson, 2003; Kuhl et al., 2007). If so, the amount of signal reduction should predict forgetting, and this relationship should arise only when reactivation occurs, as we observed. These findings are thus consistent with reactivation-induced lability (Nader et al., 2000; Dudai, 2004). Although consistent, the exact relationship to reconsolidation is unclear, as our design does not permit us to address those processes, which unfold more slowly over time.

The downregulation of hippocampal activation observed here may be analogous to downregulations of neocortical areas observed when people ignore distracting input in a visual working memory task. For instance, activation in the parahippocampal place area is reduced when places must be ignored during the encoding of faces into visual working memory (Gazzaley et al., 2005). This finding may be an instance of the principle advocated here, that purging unwanted content is achieved by suppressing neural activation supporting awareness of the purged content. What cannot be discerned in that prior work, however, is whether downregulation reflects preemptive modulation of areas representing the distracting content to prevent encoding of that information, or instead the purging of unwanted contents that have already penetrated awareness. No measures were included to differentiate trials in which distracting information entered awareness from those when it did not. Because of our use of the trial-by-trial method of introspection, the present findings are unique in tying the voluntary purging of contents that penetrated phenomenal awareness to the downregulation of neural activity supporting awareness, and, moreover, in demonstrating that such downregulation has persisting behavioral consequences consistent with inhibition.

Although the present findings point to neural activation markers that reflect the purging of mnemonic awareness, BOLD activation indirectly measures neural activity. If our mechanistic hypothesis is correct, however, one should observe in the behavior of individual neurons linked to awareness reduced firing that accompanies the intention to purge unwanted contents. Previous work has shown that ignoring an object in a visual-selective attention task is linked to downregulated activity in hippocampal neurons that preferentially respond to that object, providing an important precedent for the present hypothesis (Cerf et al., 2010). Unlike the highly targeted control response in that work, however, the present downregulation appears to be more widespread, and to have persisting consequences. Nevertheless, if these phenomena are related, purging memories from awareness may disrupt the stability of the mapping of hippocampal neurons to their preferred objects, reflecting the neural basis of memory disruption similar to that found here. Thus, observing neural activity relating to the emergence of episodic awareness and its termination by inhibitory control should be possible. Because inhibitory control is engaged in a range of cognitive domains (Anderson and Spellman, 1995; Anderson and Green, 2001; Gazzaley et al., 2005), it may also support the intentional purging of awareness in contexts other than memory. If so, the current findings provide a mechanistic hypothesis about how human beings terminate awareness of unwanted mnemonic content that may extend to the suppression of conscious content throughout the brain. The nature of phenomenal awareness itself remains unresolved (Chalmers, 1995), but the current linkage of first-person experience to the regulation of neural activity suggests that the
physical basis by which people voluntarily stop this experience may ultimately be witnessed and understood.

References


Levy and Anderson • Purging Memories from Consciousness