Two routes to emotional memory: Distinct neural processes for valence and arousal

Elizabeth A. Kensinger* and Suzanne Corkin

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129

Edited by Marcus E. Raichle, Washington University School of Medicine, St. Louis, MO, and approved December 30, 2003 (received for review October 3, 2003)

Prior investigations have demonstrated that emotional information is often better remembered than neutral information, but they have not directly contrasted effects attributable to valence and those attributable to arousal. By using functional MRI and behavioral studies, we found that distinct cognitive and neural processes contribute to emotional memory enhancement for arousing information versus valenced, nonarousing information. The former depends on an amygdalar–hippocampal network, whereas the latter was supported by a prefrontal cortex–hippocampal network implicated in controlled encoding processes. A behavioral companion study, with a divided-attention paradigm, confirmed that memory enhancement for valenced, nonarousing words relied on controlled encoding processes: concurrent task performance reduced the enhancement effect. Enhancement for arousing words occurred automatically, even when encoding resources were diverted to the secondary task.

Why do we remember some experiences while forgetting others? Neuroimaging has provided a tool to probe this question. By comparing the neural activity during encoding of items that are later remembered versus those that are later forgotten (a “subsequent memory” analysis), one can examine the processes mediating successful encoding (i.e., that are carried out during encoding of words that will later be remembered; reviewed in ref. 1). Functional MRI studies have indicated that activation in prefrontal cortex (PFC), hippocampus, and parahippocampal gyrus underlies successful encoding, such that greater activation in these regions increases the probability that information will be remembered.

In most functional MRI studies examining memory formation, stimuli are chosen deliberately to preclude an emotional response. In daily life, however, much of the information we encounter holds emotional significance. Abundant evidence indicates that we often are more likely to remember this emotional information than we are to remember information lacking in emotional import (2, 3). The goal of the present study was to uncover the neural circuits mediating this enhanced memory for emotional information.

The amygdala is an obvious site. The early studies of Kluver and Bucy (4) suggested that lesions to the amygdala can result in abnormalities in assigning emotional significance to stimuli. More recent studies have demonstrated a link between amygdalar function and explicit memory for emotional information: patients with amygdalar lesions do not show enhanced memory for emotional as compared with neutral stimuli (5–7), and neuroimaging studies have revealed a link between the amount of amygdalar activation at encoding and the likelihood of later retrieving emotional items (8–12).

Although the evidence implicating the amygdala in emotional memory is strong, further specification of the neural processes is required because the amygdala clearly does not act in isolation. Similarly, the cognitive processes that contribute to the enhancement effect require delineation. Some processes may be self-generated and controlled: individuals may be inclined to elaborate on emotional information (semantically or autobiographically) or to rehearse emotional information (see ref. 13; but see 14). For example, presentation of emotional items (e.g., coffin) may remind an individual of a personal event more often than neutral items (e.g., banner), and thus individuals may associate emotional items with personal experiences. Other processes may be relatively automatic: attention may be directed toward threatening or aversive stimuli (15–17), and these stimuli may benefit from prioritized or facilitated processing (18–20). It is reasonable to expect that different neural substrates underlie the controlled versus automatic processes that contribute to the enhancement effect. PFC has been implicated in controlled processes like elaboration or rehearsal, such that items that evoke PFC activation (and, thus, the additional use of such strategies) are more often remembered than items that engage PFC to a lesser extent (1, 21–23). In contrast, automatic capture of attention by emotion is likely mediated by the amygdala. Pharmacologic lesion studies have demonstrated that the central nucleus of the amygdala is critical for enhanced attentional arousal in response to fear-evoking stimuli (24, 25), and patient studies have confirmed that amygdalar lesions eliminate orienting toward emotional stimuli (18, 26). Amygdalar connections to the brainstem (24, 27), thalamus (24, 28), and lower-level sensory areas (28, 29) may mediate these attentional effects.

The types of processes engaged also may differ depending on the particular stimulus characteristics. Emotional information can be characterized in two dimensions: arousal (how exciting or calming) and valence (how positive or negative) (e.g., see ref. 30 and 31). Evidence suggests that the amygdala’s role in emotional memory is related to arousal (see ref. 32 for the first proposal of a relation between arousal and amygdalar engagement). Pharmacologic manipulations that increase arousal levels (e.g., administration of noradrenergic agonists) enhance memory performance in animals and humans, whereas the memory facilitation typically associated with emotional arousal is eliminated by administration of β-adrenergic antagonists (33–35). Valence and stimulus intensity also have been found to have dissociable effects in the gustatory system (36) and in the olfactory system (37), with the amygdala being particularly important for odor or taste intensity.

It is currently unclear whether similar dissociations between valence and arousal exist within memory systems. Most studies examining the effect of emotional content on long-term memory have used arousing stimuli (i.e., those high in emotional intensity); however, items that are not arousing but that have valence (and, in particular, that participants judge to be “negative”) are also better remembered than neutral stimuli (e.g., see ref. 38). The neural processes contributing to the latter enhancement effect have not been examined. Prior neuroimaging studies...
investigating the neural processes that underlie the ability to learn emotional information have used stimuli that contain valence and arousal, i.e., that are negative and arousing (29, 39, 40) or, less frequently, positive and arousing (11, 12, 41, 42). These studies have not made an explicit distinction between valence and arousal. The present study, therefore, asked whether the processes that contribute to successful memory formation for arousing items also support successful encoding of valenced, nonarousing items, or whether distinct processes underlie the memory benefit for these items. The results indicate that distinct networks support memory enhancement for words with arousal versus those with only valence. Consistent with studies linking amygdalar modulation of memory to arousal (3, 43, 44), the memory benefit for arousing words was supported by an amygdalar–hippocampal network. In contrast, the memory enhancement for negative nonarousing words was mediated via a PFC–hippocampal network also implicated in encoding of neutral words (1).

Methods

Participants and Procedures. Participants provided informed consent in a manner approved by the Massachusetts Institute of Technology and Massachusetts General Hospital Institutional Review Boards; they were remunerated at $25 per hour for their participation. Participants comprised 28 young adults (14 women and 14 men) who were scanned on a Siemens (Erlangen, Germany) Allegra 3 Tesla head-only MRI scanner while they encoded words that were neutral, negative and nonarousing (e.g., sorrow, mourning, etc.), or negative and arousing (e.g., rape, slaughter, etc.). Words were presented for 2 sec each, pseudorandomly intermixed with fixation crosses to provide jitter (45, 46), and participants rated each word as “abstract” or “concrete.”

Each encoding scan was followed by a retrieval scan (after an ~10-min delay) in which participants indicated by button press (remember, know, new) whether they (i) vividly remembered seeing the word at encoding (i.e., remembered something specific about the item’s presentation, such as a thought they had when viewing the word), (ii) sensed that the word was familiar and thus thought it had been presented at study but did not remember any details about its prior presentation (see ref. 47 for review of this distinction), or (iii) believed the word had not been presented at study. Nonstudied foils from each emotional category were included on the recognition task. Participants were aware that a recognition task would follow each encoding scan. After the encoding and retrieval sessions, participants rated the words for valence (i.e., how positive or negative) and arousal (i.e., how calming or exciting), each on a scale from 1–9. We used these ratings to place words into three categories: (i) negative (valence of −3) and nonarousing (arousal of 1–3), (ii) negative (valence of −3) and arousing (arousal of 6–9), and (iii) neutral (valence of 4–6, arousal of 1–5). On average, participants rated 125 words (SD = 24.8) as neutral, 90 (SD = 28.8) as negative and nonarousing, and 85 (SD = 46.5) as negative and arousing.

Data Analysis. We preprocessed the data by using SPM99 (Wellcome Department of Cognitive Neurology, London), correcting the images for slice timing and rigid body motion. Functional data then were normalized spatially to the Montreal Neurological Institute template. Images were resampled into 3-mm cubic voxels and smoothed spatially with an 8-mm full-width half-maximum isotropic Gaussian kernel.

Statistical analyses used the general linear model in SPM99. Trials from each condition were modeled by using a canonical hemodynamic response function. Effects for each condition were estimated by using a subject-specific, fixed-effects model. These data then were entered into a second-order, random-effects analysis. Encoding analyses contrasted the experimental trials (collapsing across item type) to the baseline (fixation). Activation was considered reliable if the area included at least 5 voxels at P < 0.001 uncorrected. The encoding-related regions of interest (ROIs) were defined from this contrast. These regions were unbiased with respect to item type and allowed us to assess the effect of emotion and subsequent memory (i.e., activation related to whether an item was remembered or forgotten) in regions associated with episodic encoding. ROIs were 8-mm spheres, except for the amygdala, where a 3-mm sphere defined the ROI.

Results

Behavioral Results. An ANOVA indicated a significant effect of emotion type (arousing, negative nonarousing, neutral) and response type (remember, know) and an interaction between emotion type and response type (all P values < 0.01). Subsequent t tests indicated that participants remembered more negative, arousing words (87%) and negative, nonarousing words (85%) than neutral words (77%). The majority of correct responses were “remember” responses (69% for negative, arousing; 67% for negative, nonarousing; and 54% for neutral); the “know” responses did not differ across the three emotional categories (18% for negative, arousing; 21% for negative, nonarousing; and 23% for neutral). These patterns also held when corrected recognition scores (hits – false alarms) were computed.

Neuroimaging Results: Encoding. A random-effects, voxel-based analysis compared brain activity during all encoding trials, collapsing across word types, as compared with fixation (Table 1). We then used this contrast to define ROIs. ROIs were created for all regions that survived the P < 0.001 threshold with a 5 voxel extent (Table 1). For each ROI, we examined the peak percentage signal change that occurred 2–6 sec after stimulus onset. An ANOVA was conducted on these signal change values

<p>| Table 1. Regions activated during encoding (at least 5 voxels, P &lt; 0.001 uncorrected) |
|-----------------------------|-----------------|---------------|</p>
<table>
<thead>
<tr>
<th>Brain region</th>
<th>Hemisphere</th>
<th>Talairach coordinates, x, y, z</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occipital lobe</td>
<td>L</td>
<td>−21, −102, 0</td>
<td>17/18</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>27, −99, −3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>33, −93, −6</td>
<td></td>
</tr>
<tr>
<td>Inferior parietal lobe</td>
<td>L</td>
<td>−48, −42, 54</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>54, −33, 51</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>−54, 24, −9</td>
<td>47</td>
</tr>
<tr>
<td>Inferior prefrontal gyrus</td>
<td>L</td>
<td>−51, 36, 15</td>
<td>45/46</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>−51, 18, 30</td>
<td>9/44</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>48, 24, −15</td>
<td>47</td>
</tr>
<tr>
<td>dorsolateral PFC</td>
<td>L</td>
<td>−48, 33, 27</td>
<td>9/46</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>48, 42, 15</td>
<td>9/46</td>
</tr>
<tr>
<td>Superior prefrontal gyrus</td>
<td>L</td>
<td>−3, 27, 42</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>−6, 18, 66</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>−30, 9, 60</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>0, 18, 51</td>
<td>8</td>
</tr>
<tr>
<td>Uncal hippocampus</td>
<td>R</td>
<td>27, 6, −18</td>
<td></td>
</tr>
<tr>
<td>Anterior hippocampus</td>
<td>L</td>
<td>−30, −15, −12</td>
<td></td>
</tr>
<tr>
<td>Amygdala</td>
<td>L</td>
<td>−27, −3, −12</td>
<td></td>
</tr>
<tr>
<td>Claustrum</td>
<td>R</td>
<td>27, 0, −18</td>
<td></td>
</tr>
<tr>
<td>Striatum</td>
<td>L</td>
<td>−24, −6, 6</td>
<td></td>
</tr>
</tbody>
</table>

L, left; R, right.

*This ROI size was chosen because the amygdala is a punctate region and larger spheres included activation beyond the amygdala proper.*
to examine the effects of emotion type (negative arousing, negative nonarousing, neutral) and subsequent memory (remembered, forgotten). Below we report results from the ROIs that demonstrated each effect.

**Emotion Effect.** In the left hemisphere, the hippocampus, amygdala, and inferior parietal lobule [Brodmann’s area (BA) 40] showed a greater response to negative arousing and negative nonarousing words than to neutral words (Fig. 1). Thus, these regions were modulated by the presence of any emotional salience (valence or arousal).

The inferior parietal lobule has been implicated in processing of verbal information related to the self (49), attention (see ref. 50 for review), and working memory processing of emotional content (51). Any of these possibilities could explain the recruitment of inferior parietal lobule for processing emotional categories of words.

The modulation of the amygdala to the negative nonarousing words suggests that, with verbal stimuli at least, the amygdala may not be selectively modulated by arousal. In the present study, the amygdala showed above-baseline activation even for the neutral words. It has been proposed that the amygdala may be engaged during the processing of ambiguous stimuli (52, 53). The amygdala, therefore, may have been engaged during the processing of all of the words because the verbal stimuli all had to be processed to some level before their threat (or lack thereof) could be evaluated.

In contrast to these regions that were modulated by any emotional salience (arousing or negative nonarousing words), the left inferior PFC (BA 47) and left dorsolateral PFC (BA 9 and 46) showed greatest activation during encoding of negative nonarousing words (i.e., valence-only words) as compared with arousing or neutral words (Fig. 2). The activation in the left PFC may reflect additional, self-initiated encoding processes that were carried out on the valence-only words, such as autobiographical or semantic elaboration or additional rehearsal. This explanation is consistent with prior studies implicating these PFC regions with elaborative encoding processes (21–23, 54).

**Subsequent Memory Effect.** To gain leverage on whether distinct neural processes contributed to the successful encoding of arousing words versus negative nonarousing words, encoding-related brain activity in the defined ROIs was compared for words that were later vividly remembered by the participant as compared with words that were later forgotten (i.e., participants later indicated incorrectly that the word had not been previously studied). Because accuracy for the arousing words was very high, the subsequent memory analyses are shown for the 19 individuals (9 women and 10 men) who had a sufficient number of forgotten, arousing words (at least 12) to permit the subsequent memory analysis. When the data from all 28 participants were analyzed, the results for the negative nonarousing and neutral words remained qualitatively the same as those for the 19 individuals.

This subsequent memory analysis revealed distinct networks supporting memory formation for arousing words versus nonarousing (negative nonarousing and neutral) words. Activation in the left amygdala and left hippocampus correlated with successful memory formation for the arousing words (Fig. 3). Furthermore, we found a significant correlation between the activation in these two regions during the encoding of subsequently remembered, arousing words (Fig. 4).

We found a second network in the left inferior PFC and left hippocampus related to successful memory formation (greater activation during the encoding of subsequently remembered versus subsequently forgotten words) for the negative nonarous-
ing and neutral words (Fig. 3). An ANOVA with region, emotion type, and subsequent memory as factors indicated a significant interaction between item type and subsequent memory ($P < 0.01$). Activation in these regions was stronger for subsequently remembered negative nonarousing words than for subsequently remembered neutral words, whereas the activation to forgotten neutral words did not differ.

These results suggest that distinct cognitive and neural processes contribute to memory formation for arousing words versus words without arousal (negative nonarousing and neutral). Specifically, amygdalar activation correlated with subsequent memory performance only when the items were arousing. This result is consistent with prior studies in which amygdalar activation related to successful memory formation only for items that participants rated as highly emotional (9). Furthermore, for subsequently remembered arousing words, percentage signal change in the amygdala and hippocampus was correlated (see also ref. 12), consistent with the hypothesis that amygdalar–hippocampal interactions are critical for the emotional memory enhancement effect (35).

In contrast to the amygdalar–hippocampal network recruited for encoding of arousing words, left inferior PFC activation supported successful memory formation for negative nonarousing words, as well as for words that were neutral, but not for arousing words. Furthermore, the greater activation in this region during the successful encoding of negative nonarousing words as compared with neutral words suggests that negative nonarousing items show memory benefits because of additional engagement of controlled cognitive processes (e.g., elaboration) supported by the left inferior PFC.

**Behavioral Companion Study.** The subsequent memory analyses supported the hypothesis that amygdalar modulation of hippocampal function underlies the memory benefit for arousing words (35). Amygdalar activation to threatening or aversive stimuli is believed to occur relatively automatically and even when attentional resources are taxed (16, 20). Thus, amygdalar modulation of hippocampal function could occur relatively automatically. In contrast, the neuroimaging data reported here suggest that the processes associated with the enhancement for the negative nonarousing items are attention-demanding, controlled processes, similar to those engaged during the encoding of neutral words.

To seek support for this interpretation, we conducted a
behavioral study by using a divided-attention paradigm. The question asked was whether additional encoding resources were required for the memory enhancement for negative nonarousing words versus arousing words. In the divided-attention paradigm, participants perform a concurrent task as they encode words. By varying the difficulty of this concurrent task, we could modulate the resources available for encoding (e.g., see ref. 55). This manipulation has been shown to have a greater impact on controlled encoding processes and a lesser effect on relatively automatic processes (47). Thus, if self-generated, controlled encoding processes were responsible for the enhancement effect for negative nonarousing items, but were less important for the enhancement for arousing items, the enhancement for the negative nonarousing items should be disproportionately reduced by the divided-attention manipulation.

To test this hypothesis, 24 participants (12 women and 12 men) encoded words used in the functional MRI experiment while (i) performing a hard auditory discrimination task, (ii) performing an easy auditory discrimination task, or (iii) performing no secondary task (see ref. 56 for details of the auditory discrimination task). The results supported the hypothesis that encoding of negative nonarousing words disproportionately recruited self-generated encoding processes. With no concurrent task, there was a benefit for the negative arousing words (88%, SE = 3%) and the negative nonarousing words (83%, SE = 3%) as compared with the neutral words (75%, SE = 3%). When individuals performed the easy concurrent task, however, they showed no memory enhancement for the negative nonarousing words (71%, SE = 3%) as compared with the neutral words (68%, SE = 3%) but did show significantly better memory for the arousing words (84%, SE = 3%). These results were replicated with the difficult secondary task. There was no difference in memory for the negative nonarousing words (62%, SE = 3%) as compared with the neutral words (60%, SE = 3%) but significant enhancement for the arousing words (83%, SE = 3%).

**Discussion**

Two main conclusions emerge from the present study. First, we report evidence in humans of a link among amygdalar activation, hippocampal activation, and subsequent memory (see also ref. 12). During successful encoding of arousing words, the activation in these regions was correlated. This correlation is consistent with the hypothesis (34) that activation in the amygdala results in modulation of hippocampal function (although the correlation cannot provide evidence of the direction of modulation, i.e., amygdalar→hippocampal or hippocampal→amygdalar). The fact that this relation occurred only for arousing words also is consistent with the modulation hypothesis. Although the details of the modulatory effect are still being uncovered, it likely results from effects of stress hormones (epinephrine and corticosteroids, which are released as part of the response to emotionally arousing events) on the limbic system (35). The results of our behavioral companion study further suggest that this modulation may occur even when attentional resources are taxed. Thus, even when encoding resources were devoted toward a secondary task, the memory enhancement for arousing words remained. This result corroborates evidence suggesting that amygdalar activation by emotional stimuli can occur even with limited attention (refs. 16 and 20; see ref. 57 for evidence that some attention to emotional stimuli may be necessary) and that at least some attention-demanding processes, such as increased rehearsal, are not sufficient to explain the memory enhancement for emotional information (14).

Second, this study indicates that although arousing words and negative nonarousing words are, at least in some instances, more likely to be remembered than neutral words (38), distinct cognitive and neural processes contribute to these enhancement effects. Memory for negative nonarousing words is enhanced because of additional recruitment of the same types of self-generated, controlled processes as are used to encode neutral words. Individuals may be more likely to elaborate on, or to rehearse, these negative nonarousing words as compared with neutral words. This hypothesis is supported by the neuroimaging data. Activation in regions associated with controlled encoding processes (e.g., PFC) was greater for valence-only words than for neutral words, and the relation between activation and subsequent memory in these regions was stronger for valence-only words than for neutral words. The results from the divided-attention behavioral companion study supported this hypothesis. Diversion of resources from the encoding task eliminated the memory enhancement for the negative nonarousing words, presumably by reducing the participants’ ability to engage in these additional, controlled processes (e.g., ref. 55).

In summary, we propose that two distinct mechanisms support memory enhancement for emotional information, depending on whether that information is arousing or negative but not arousing. The enhancement for negative nonarousing items is supported by a PFC→hippocampal network that has been implicated in memory formation for neutral information (e.g., see ref. 1) and is associated with controlled, self-generated encoding processes, such as elaboration or rehearsal of information (e.g., see refs. 54 and 56). In contrast, memory enhancement for arousing items is mediated by an amygdalar→hippocampal network, which may reflect relatively automatic effects of emotion on memory, and may be specifically engaged when emotional stimuli elicit an arousal effect.

We thank Anthony Wagner, Helen Barbas, Earl Miller, Nancy Kanwisher, Anne Krendl, Gail O’Kane, and Olivier Piguet for helpful discussion and comments on an earlier version of this manuscript. This research was supported by National Institute of Health Grant AG021525, National Science Foundation Doctoral Dissertation Research Grant BCS0212999, National Center for Research Resources Grant P41RR14075, the Mental Illness and Neuroscience Discovery (MIND) Institute, and a Howard Hughes Medical Institute Predoctoral Fellowship (to E.A.K).

---