Emotional arousal impairs association memory: roles of prefrontal cortex regions

Esther Fujiwara,1,2 Christopher R. Madan,1,3 Jeremy B. Caplan,1,2 and Tobias Sommer1

1University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany; 2University of Alberta, Edmonton, Alberta T6G 2R3, Canada; 3University of Nottingham, Nottingham NG5 1PB, United Kingdom

The brain processes underlying impairing effects of emotional arousal on associative memory were previously attributed to two dissociable routes using high-resolution fMRI of the MTL (Madan et al. 2017). Extrahippocampal MTL regions supporting associative encoding of neutral pairs suggested unitization; conversely, associative encoding of negative pairs involved compensatory hippocampal activity. Here, whole-brain fMRI revealed prefrontal contributions: dmPFC was more involved in hippocampal-dependent negative pair learning and vmPFC in extrahippocampal neutral pair learning. Successful encoding of emotional memory associations may require emotion regulation/conflict resolution (dmPFC), while neutral memory associations may be accomplished by anchoring new information to prior knowledge (vmPFC).

Emotional arousal is well known to enhance memory for individual items (Schümann and Sommer 2018), but can have impairing effects on associative memory, particularly when items cannot be easily unitized and interactional associations have to be formed (Madan et al. 2012; Murray and Kensinger 2013; Bisby and Burgess 2017). The neural substrates of the impairing effect of emotional arousal on associative memory have only begun to be explored (Bisby et al. 2016; Madan et al. 2017). Emotional arousal may disrupt hippocampal functions that are critical to promote binding and thereby lead to reduced associative memory for emotionally arousing items (“disruption hypothesis”) (Bisby et al. 2016). Conversely, encoding of neutral items may engage extrahippocampal medial temporal lobe (MTL) regions, areas we interpreted to promote better incidental unitization of neutral than negative items, leading to a net-decrease in associative memory for negative items (“bypassing hypothesis”) (Madan et al. 2017).

Specifically, in our previous high-resolution fMRI study focusing on MTL regions (Madan et al. 2017), extrahippocampal MTL cortex was more active during encoding of neutral than negative picture pairs, showed a subsequent memory effect (SME) for neutral picture pairs, and neutral pair encoding was accompanied by more between-picture saccadic eye movements compared with negative pairs. In line with previous findings of extrahippocampal MTL areas involved in association memory formation of merged or unitized items (Giovanello et al. 2006; Quamme et al. 2007; Diana et al. 2008; Delhaye et al. 2019), we interpreted our fMRI and eye movement findings to suggest better incidental unitization of neutral picture pairs than negative pictures. A behavioral follow-up study confirmed that unitization, that is, imagining the two pictures as one (“interactive imagery”), was indeed rated as higher for neutral than negative pairs, and this advantage in interactive imagery was linked to better associative memory for neutral pairs, lending further support to the bypassing hypothesis (Caplan et al. 2019).

What would prevent emotional pairs from being as easily merged as neutral pairs? We observed that during negative pair encoding, each individual picture was fixated longer compared with neutral pictures. These longer fixation durations for negative pictures were related to greater activity during negative than neutral pair encoding in the dorsal amygdala (likely the centromedial group) (Hrybouski et al. 2016), an activation which was functionally coupled to the more ventral amygdala (likely the lateral nucleus) (Hrybouski et al. 2016). This ventral amygdala activation exhibited a subsequent forgetting effect specifically for negative pairs. Given that emotional items attract more attention to themselves and are more likely processed as individual items (Markovic et al. 2014; Mathé et al. 2016), we conjectured that this may make pairs of emotional items harder to unitize and to benefit from extrahippocampal unitization-related processes such as interactive imagery. Interestingly, the hippocampus revealed a subsequent memory effect specifically for negative pairs in Madan et al. (2017). We concluded that when sufficiently arousing information precludes extrahippocampal unitization-based encoding, an alternative, compensatory, and effortful relational hippocampus-dependent encoding strategy may be used.

Both findings, extrahippocampal MTL encoding for neutral pairs and compensatory hippocampal encoding for negative pairs, raise the question of which cortical areas could be involved in these two dissociable associative encoding processes. Conceivably, successful associative encoding of negative information may require participants to evaluate the emotional content, and regulate emotional arousal/conflict, functions primarily associated with dorso-medial PFC regions (dmPFC; Dixon et al. 2017), the anterior cingulate cortex (ACC) (Botvinick 2007), and posterior areas of the ventro-medial PFC (vmPFC) (Yang et al. 2020). To unitize two pictures through interactive imagery, retrieval of semantic memories and prior knowledge regarding the contents of the two pictures is likely helpful. Semantic memory processes have been attributed to the left inferior frontal gyrus (left IFG) (Binder and Desai 2011). The vmPFC (more anterior portions) could also be involved, owing to its role in relating new information during encoding to prior knowledge, that is, a “unitization-like” process (Gilboa and Marlatte 2017; Sommer 2017). Motivated by our previous discovery and interpretation of the two distinct encoding processes in the MTL (Madan et al. 2017), the potential contribution of these cortical areas in neutral and negative association memory was...
explore here by using a whole-brain scan and overcoming the limitations of our previous high-resolution fMRI sequence focused only on the MTL in Madan et al. (2017).

In the current study, we therefore acquired standard-resolution whole-brain fMRI (3 Tesla Siemens Trio scanner, 3-mm thickness, TR 2.21 sec, TE 30 msec) of 22 male participants during exactly the same task as in Madan et al. (2017). Only male participants were recruited because of known sex-dependent lateralization of amygdala activity (Cahill et al. 2004; Mackiewicz et al. 2006), limiting the conclusions of the current study to males.

Eye movements were assessed as a proxy for overt attention (EyeLink 1000, SR Research, 17 participants with usable eye-tracking data). In each of three encoding-retrieval cycles, 25 neutral and 25 negative picture pairs were intentionally encoded. Pictures (e.g., objects, scenes, humans, and animals) were selected from the International Affective Picture System (Lang et al. 2008) and the internet, and confirmed to have different valence and arousal through independent raters (details in Madan et al. 2017). Each encoding round was followed by a two-step memory test for each pair: In a judgement of memory (JoM) task one picture served as retrieval cue and volunteers indicated their memory (yes/no) for the other picture of the original pair. Then the same picture served as retrieval cue and volunteers indicated their memory (fixation duration and number of within-picture saccades), but this was unrelated to association memory success.

The fMRI data were preprocessed (slice timing corrected, realigned and unwarped, normalized using DARTEL and smoothed, FWHM = 8 mm) and analyzed using SPM12. First-level models were created with four regressors that modeled the onsets of the 2 (negative and neutral) × 2 (subsequent hits and misses) conditions of interest. Results of all fMRI analyses were considered significant at P < 0.05, family-wise error (FWE) corrected for multiple comparisons across the entire scan volume or within the a priori anatomical regions of interest (ROIs). ROIs for the hippocampus, amygdala and extrahippocampal MTL were reused from our previous study (Madan et al. 2017). The prefrontal ROIs, that is, dmPFC, ACC, vmPFC and left inferior frontal gyrus ROIs were manually traced on the mean T1 image using ITK-SNAP 3.6.0 (Yushkevich et al. 2006) following schematic drawings based on meta-analyses (Binder and Desai 2011; Dixon et al. 2017; Gilboa and Marlatte 2017).

The second-level analyses based on the resulting individual β images and subject as a random factor replicated a well-established network of brain areas involved in negative emotion processing (Spalek et al. 2015): greater activity during processing negative than neutral pictures in the amygdala, insula, right inferior frontal gyrus, mid, and anterior cingulate cortex as well as visual areas (Fig. 2A). As in our previous study, we correlated the difference in left amygdala activity with the difference in eye movements for negative minus neutral trials, showing a significant correlation with the number of within-picture saccades (r = 0.50, P = 0.018). Thus, higher left amygdala activity was associated with increased visual search within negative pictures. We conducted a psychophysiological interaction analysis (PPI) using this amygdala region as seed and contrasted functional coupling during successful versus unsuccessful negative with successful versus unsuccessful neutral pair encoding (i.e., the interaction of valence and subsequent memory success). This PPI revealed stronger coupling during successful encoding of negative compared with neutral pairs with a (nonsignificant) cluster in the dmPFC (Z = 3.01, [−12, 38, 26]). Simple effects showed that the amygdala was more strongly coupled with the dmPFC during successful than unsuccessful negative pair encoding (Z = 3.63, [−2, 16, 42]).

Neutral-pair processing was associated with greater activity than negative-pair processing in the bilateral extrahippocampal MTL cortex, ventral precentral (vPC), retrosplenial cortex (RSC), middle occipital gyrus, and putamen (Fig. 2B). In addition, we observed a general SME irrespective of valence in the left hippocampus (−28, −16, −24), Z = 3.49, P = 0.04).

An interaction between pair valence and SME with greater neutral than negative SME was observed in vmPFC (Fig. 3A), together with a (nonsignificant) cluster in right MTL cortex (I26, −24, −28), Z = 3.16, P = 0.011. We conducted a PPI using this vmPFC region as seed and contrasted...
The current findings, first, replicated the impairing effects of emotional arousal on association memory previously observed in six experiments across four studies (Madan et al. 2012, 2017; Caplan et al. 2019). We built on these previous findings here by identifying cortical, especially prefrontal areas involved in the associative memory advantage for neutral pairs and those involved in the compensatory mechanism for learning negative pairs. In particular, vmPFC activity more strongly supported successful encoding of neutral than negative pairs and during this process, showed stronger coupling with a cluster at the border between MTL cortex and hippocampus. Conversely, the dmPFC was more engaged and more strongly coupled with the hippocampus during successful negative than neutral pair encoding.

We observed more and longer fixations, as well as more within-picture saccades for individual negative pictures compared with neutral pictures, resembling previously reported eye movement findings (Bradley et al. 2011; Dietz et al. 2011). We had previously shown that increased attention (fixation duration) to individual negative pictures is linked to centromedial amygdala activity (not measurable here due to the whole-brain scan resolution), and functionally coupled with a negative pair-specific subsequent forgetting effect in the lateral amygdala (Madan et al. 2017). These findings together suggest that increased attention attracted by individual negative pictures does not support associative memory, or may even be detrimental (cf., Hockley and Cristi 1996).

Figure 2. Main effects of emotion—fMRI results. (A) Greater activity during negative than neutral pair processing irrespective of subsequent memory success. (B) Greater activity during neutral than negative pairs processing. t-maps thresholded at $P<0.001$ uncorrected for visualization purposes. t-value color-coded.

Figure 3. SME × Emotion interactions and PPIs. (A) Activity in the vmPFC revealed a SME only for neutral but not negative pairs. (B) This region was stronger coupled during neutral than negative pair encoding with a cluster in the border of left MTL cortex/hippocampus. (C,D) Activity in the right hippocampus and dmPFC revealed a SME only for negative pairs. (E) The dmPFC was stronger coupled during negative than neutral pairs encoding with the bilateral hippocampus. t-maps thresholded at $P<0.001$ uncorrected for visualization purposes. Error bars are 95% confidence intervals around the mean, corrected for interindividual differences (Loftus and Mason 1994).

The current findings, first, replicated the impairing effects of emotional arousal on association memory previously observed in six experiments across four studies (Madan et al. 2012, 2017; Caplan et al. 2019). We built on these previous findings here by identifying cortical, especially prefrontal areas involved in the associative memory advantage for neutral pairs and those involved in the compensatory mechanism for learning negative pairs. In particular, vmPFC activity more strongly supported successful encoding of neutral than negative pairs and during this process, showed stronger coupling with a cluster at the border between MTL cortex and hippocampus. Conversely, the dmPFC was more engaged and more strongly coupled with the hippocampus during successful negative than neutral pair encoding.

We observed more and longer fixations, as well as more within-picture saccades for individual negative pictures compared with neutral pictures, resembling previously reported eye movement findings (Bradley et al. 2011; Dietz et al. 2011). We had previously shown that increased attention (fixation duration) to individual negative pictures is linked to centromedial amygdala activity (not measurable here due to the whole-brain scan resolution), and functionally coupled with a negative pair-specific subsequent forgetting effect in the lateral amygdala (Madan et al. 2017). These findings together suggest that increased attention attracted by individual negative pictures does not support associative memory, or may even be detrimental (cf., Hockley and Cristi 1996).

The current findings, first, replicated the impairing effects of emotional arousal on association memory previously observed in six experiments across four studies (Madan et al. 2012, 2017; Caplan et al. 2019). We built on these previous findings here by identifying cortical, especially prefrontal areas involved in the associative memory advantage for neutral pairs and those involved in the compensatory mechanism for learning negative pairs. In particular, vmPFC activity more strongly supported successful encoding of neutral than negative pairs and during this process, showed stronger coupling with a cluster at the border between MTL cortex and hippocampus. Conversely, the dmPFC was more engaged and more strongly coupled with the hippocampus during successful negative than neutral pair encoding.

We observed more and longer fixations, as well as more within-picture saccades for individual negative pictures compared with neutral pictures, resembling previously reported eye movement findings (Bradley et al. 2011; Dietz et al. 2011). We had previously shown that increased attention (fixation duration) to individual negative pictures is linked to centromedial amygdala activity (not measurable here due to the whole-brain scan resolution), and functionally coupled with a negative pair-specific subsequent forgetting effect in the lateral amygdala (Madan et al. 2017). These findings together suggest that increased attention attracted by individual negative pictures does not support associative memory, or may even be detrimental (cf., Hockley and Cristi 1996).
The dmPFC contributed more to negative than neutral associative memory and was functionally coupled to the hippocampus, which complements our interpretation of possibly compensatory activity in the hippocampus during negative pair encoding (Madan et al. 2017). The amygdala on the other hand was stronger activity in the hippocampus during negative pair encoding which complements our interpretation of possibly compensatory memory and was functionally coupled to the hippocampus, with the bilateral hippocampus during successful negative compared with neutral pair learning. The involvement of the dmPFC during successful negative (but unsuccessful neutral) (discussed below) pair encoding may suggest that it resolves conflicts between the prepotent attention to the individual negative pictures and the current task goals, that is, their intentional associative encoding. One way to do so might involve the dmPFC’s role to regulate the negative emotions elicited by the pictures in order to focus on the associative memory task.

Neutral pairs elicited more between-picture saccades than negative pairs, as in (Madan et al. 2017). The vmPFC was more strongly involved in successful associative encoding of neutral than negative pairs and more strongly coupled with the extrahippocampal MTL cortex bordering the hippocampus during successful neutral compared with negative pair encoding. Anterior vmPFC regions and their coupling with the MTL have been implicated in retrieval of consolidated memories and in anchoring new information to prior knowledge (Nieuwenhuis and Takashima 2011; van Kesteren et al. 2013; Schlichting and Preston 2015; Gilboa and Marlatte 2017; Sommer 2017; Brod and Shing 2018; Sekeres et al. 2018). We previously observed that interactive imagery (forming one instead of two images to memorize) was higher for neutral than negative pairs (Caplan et al. 2019), perhaps reflected by the increased between-picture saccades in the current study. Assuming that the anterior vmPFC subserves retrieval of prior knowledge, its engagement during successful neutral pair encoding may have supported such incidental unitization processes here as well. Negative pictures are inherently semantically more related (Barnacle et al. 2016), which implies that they may share even more prior knowledge than neutral pictures. However, the retrieval of this prior knowledge may be inhibited by the attraction of attention to individual negative pictures, not their arbitrary pairing as in the current task. Incidental unitization can occur through rather subtle manipulations (Giovanello et al. 2006; Diana et al. 2008; Bader et al. 2010; Ford et al. 2010; Li et al. 2019) or even entirely without any instruction; for example, when the items’ combination is itself meaningful or familiar (Ahmad and Hockley 2014). We suggest that similar incidental unitization processes may have occurred here as well. Memory for unitized associations is independent of hippocampal memory processes and can be based solely on the extrahippocampal MTL (Quamme et al. 2007; Haskins et al. 2008; Staresina and Davachi 2015). Our previous high-resolution fMRI study supported such a bypassing hypothesis, that is, extrahippocampal MTL cortex involvement in the successful associative encoding of neutral but not negative pairs (Madan et al. 2017). Here, this interaction did not reach significance in the MTL cortex, but the P-value of 0.11 can be considered suggestive based on our strong a priori hypothesis. Notably, in our previous study using a scanning resolution of 1 mm$^3$ the cluster included only 17 voxels, which would correspond to less than one voxel here. Therefore, we assume the lower sensitivity here was due to the lower spatial resolution.

Unexpectedly, we observed greater activity during unsuccessful encoding of neutral pairs in the same regions that promoted successful encoding of negative pairs, that is, the dmPFC and hippocampal region. Hockley et al. (2016) previously observed that incidental but not intentional encoding of associations (for word pairs) improved for items with stronger pre-experimental associations. Perhaps using an effortful (dmPFC/hippocampal) learning strategy for neutral pairs, that is, pairs that are already more likely incidentally linked or linkable (e.g., through interactive imagery) may not have helped encoding. The forgotten neutral pairs underlying the SFE in these regions may then have been simply the hardest-to-learn neutral pairs; that is, pairs where both encoding strategies failed. Evidently, future studies should test such speculations directly.

Our interpretation of the dmPFC and vmPFC as signifying in emotion regulation and unitization in this task was based on previous studies. Because we did not manipulate unitization and/or emotional regulation, these processes remain hypothetical. However, within this framework, we addressed two hypotheses regarding interactions between hippocampal/extrahippocampal MTL regions and prefrontal cortex during association memory formation. The disruption hypothesis proposes that the hippocampus is equally responsible for encoding of negative and neutral association memory but that for negative memories, hippocampal activity is inhibited by the amygdala via the prefrontal cortex (Murray and Kensinger 2013; Bisby et al. 2016). The vmPFC has known involvement in negative emotion processing (Yang et al. 2020), and the observed activity pattern in the vmPFC could appear to disrupt hippocampal association memory processes for negative pairs. However, according to the bypassing hypothesis (Madan et al. 2017), successful encoding of negative (compared with neutral) pairs requires the hippocampus since fewer extrahippocampal contributions are available. Supporting the bypassing hypothesis, we observed that the vmPFC was negatively functionally coupled with extrahippocampal MTL cortex (bordering the hippocampus), suggesting that the vmPFC decreased extrahippocampal contributions to association memory for negative pairs. The bypassing hypothesis is also supported by our finding that the hippocampus was not less but more involved in negative compared with neutral pair encoding, that is, we observed no evidence for a prefrontally (e.g., vmPFC)-mediated disruption of hippocampal activity by emotion.

In conclusion, the two critical prefrontal cortex regions linked to MTL memory processes in the current study were the dmPFC, involved in successful hippocampal-dependent negative pair learning and the vmPFC, supporting successful neutral pair learning that relied on extrahippocampal MTL involvement.

Acknowledgments

This research was supported by a scholarship from the Deutscher Akademischer Austauschdienst (German Academic Exchange Service) to C.R.M. We also acknowledge support from the Natural Sciences and Engineering Research Council of Canada.

References

Ahmad F, Hockley WE. 2014. The role of familiarity in associative recognition of unitized compound word pairs. Q J Exp Psychol 67: 2301–2324. doi:10.1080/17470218.2014.923007

Bader R, Mecklinger A, Hoppstädter M, Meyer P. 2010. Recognition memory for one-trial-unitized word pairs: evidence from event-related potentials. Neuroimage 50: 772–781. doi:10.1016/j.neuroimage.2009.12.100
Weissman DH. 2004. Dorsal anterior cingulate cortex resolves conflict from distracting stimuli by boosting attention toward relevant events. *Cereb Cortex* **15**: 229–237. doi:10.1093/cercor/bhh125

Yang M, Tsai S-J, Li C-SR. 2020. Concurrent amygdalar and ventromedial prefrontal cortical responses during emotion processing: a meta-analysis of the effects of valence of emotion and passive exposure versus active regulation. *Brain Struct Funct* **225**: 345–363. doi:10.1007/s00429-019-02007-3

Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**: 1116–1128. doi:10.1016/j.neuroimage.2006.01.015

Received August 18, 2020; accepted in revised form November 13, 2020.