Emotional arousal impairs association-memory: Roles of amygdala and hippocampus

Christopher R. Madan[1,2,3] *, Esther Fujiwara [2,1] *, Jeremy B. Caplan [2,1] , Tobias Sommer[1]

[1] University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2] University of Alberta, Edmonton, AB, Canada
[3] Boston College, Chestnut Hill, MA, USA

* Both authors contributed equally.

2 Tables

5 Figures

Correspondence to:
Tobias Sommer
Department of Systems Neuroscience
University Medical Center Hamburg-Eppendorf, Bldg. W34
Martinistr. 52
20246 Hamburg, Germany
email: tsommer@uke.de
fon: 0049-40-7410-54763
fax: 0049-40-7410-59955

Abstract

Emotional arousal is well-known to enhance memory for individual items or events, whereas it can impair association memory. The neural mechanism of an association memory impairment by emotion is not known: In response to emotionally arousing information, amygdala activity may interfere with hippocampal associative encoding (e.g., via prefrontal cortex). Alternatively, emotional information may be harder to unitize, resulting in reduced availability of extra-hippocampal medial temporal lobe support for emotional than neutral association-memory. To test these opposing hypotheses, we compared neural processes underlying successful and unsuccessful encoding of emotional and neutral associations. Participants intentionally studied pairs of neutral and negative pictures (Experiments 1–3). We found reduced association-memory for negative pictures in all experiments, accompanied by item-memory increases in Experiment 2. High-resolution fMRI (Experiment 3) indicated that reductions in associative encoding of emotional information are localizable to an area in ventral-lateral amygdala, driven by attentional/salience effects in the central amygdala. Hippocampal activity was similar during both pair types, but a left hippocampal cluster related to successful encoding was observed only for negative pairs. Extra-hippocampal associative memory processes (e.g., unitization) were more effective for neutral than emotional materials. Our findings suggest that reduced emotional association memory is accompanied by increases in activity and functional coupling within the amygdala. This did not disrupt hippocampal association-memory processes, which indeed were critical for successful emotional association memory formation.

Significance Statement

1. Association-memory for emotional items is often worse than for neutral items.
2. This has been proposed to result from the amygdala disrupting hippocampal function.
3. We found evidence for parallel, not opposing, roles of amygdala and hippocampus.
4. Forgetting of emotional associations is driven by the amygdala.
5. But successful encoding of emotional associations continues to engage the hippocampus.

1. Introduction

Emotional arousal enhances memory for individual items or events, a robust and intensely characterized effect that generalizes across many materials and paradigms (Bradley et al., 1992; Brown and Kulik, 1977; Cahill and McGaugh, 1998). Effects of emotional arousal on association-memory are more controversial, including null-effects, increases and decreases (reviews: Mather, 2007; Mather and Sutherland, 2011; Murray and Kensinger, 2013; Yonelinas and Ritchey, 2015). Emotional arousal may enhance associative memory when the associated information can be merged so that it effectively functions like one item, e.g., the font color of a negative word or an object in placed in a semantically relevant scene (D'Argembeau and Van der Linden, 2004; Kensinger and Corkin, 2003; Mickley Steinmetz et al., 2016). In this view, the sometimes-observed enhancement of emotional associative memory may be due to the same memory-enhancing mechanism that operates on emotional items. However, if to-be-associated information cannot be easily unitized (Pierce and Kensinger, 2011; Rimmele et al., 2011) and inter-item associations have to be formed, then emotional arousal often impairs associative memory (Mather, 2007; Murray and Kensinger, 2013). These opposing but presumably simultaneous effects of emotional arousal on item-memory and inter-item associations have been recently demonstrated in the same experiment. Using a verbal associative memory paradigm, Madan et al. (2012) showed, experimentally and with mathematical modeling, that emotional arousal enhanced memory for individual emotional items (words) and simultaneously impaired associative binding between items. These results were confirmed with pairs of pictures instead of words (Bisby and Burgess, 2014; Bisby et al., 2016).

Whereas the neural processes underlying the enhancing effects of emotional arousal on item memory have been intensely characterized (Dolcos et al., 2012; Murty et al., 2010), the

neural substrates of the impairing effect of emotional arousal on associative memory have only begun to be explored (Berkers et al., 2016; Bisby et al., 2016; Murray and Kensinger, 2014). Here we adapted Madan et al.'s (2012) paradigm for the use with fMRI, a procedure that had produced simultaneous item-memory enhancing and association-memory impairing effects of emotional arousal. Our task was designed to equalize attention within and across pairs by having the two elements of the association be of the same kind (picture-picture pairs) and same valence within a given pair, and by using an intentional associative encoding instruction. Our goal was to elucidate the neural substrates of emotional versus neutral associative memory formation by focusing on the amygdala, hippocampal and MTL-cortex regions. In relation to previous neuroimaging studies, several complications in their tasks used to assess emotional association-memory are addressed with our paradigm. First, emotionally arousing information will inevitably draw or hold attention. Mixing arousing with non-arousing information in association memory studies will exaggerate this effect. Bisby et al. (2016) were the only fMRI study using pure picture pairs. Secondly, a further complication is the combination of different types of information within an association (e.g., face-occupation pairings in Berkers et al. 2016; adjective-face pairings in Okada et al., 2011), which alone could have different attentional demands (see also the relevant source memory studies: Dougal et al., 2007; Kensinger and Schacter, 2006a) where sources were always neutral and of a different kind than the items). Finally, the predominant use of incidental encoding instructions cannot address if participants attended to pair-types in the same or different way. Intentional instructions, explicitly asking participants to engage in relational encoding, should minimize attentional differences between pair-types. Although three prior fMRI studies used intentional instructions, two of these (Okada et al., 2011; Onoda et al., 2009) had a blocked fMRI design disallowing interpretation of

resulting brain activity as memory-relevant, and Berkers et al. (2016) asked participants to simultaneously perform plausibility judgements on each pair. Taken together, we think the paradigm used here can better assess the involvement in the amygdala and hippocampus in the impairment of association-memory due to emotion.

Based on the extant literature, two alternative neural mechanisms can be hypothesized that underlie better memory for neutral than emotional pairs. Both hypotheses are based on the central role of the amygdala in processing emotional arousal and in subsequent modulation of activity in other brain areas including the medial temporal lobe (MTL) (Sah et al., 2003). Both hypotheses further implicate the hippocampus and extra-hippocampal MTL-regions, given their established role in (neutral) associative and item-memory encoding (Diana et al., 2007; Eichenbaum et al., 2007). According to the first hypothesis, 'disruption hypothesis', the hippocampus remains responsible for association-memory encoding even when dealing with emotional information. As suggested by several authors, the increase in amygdala activity due to emotional arousal might lead to a disruption of hippocampus-dependent associative memory processes, reflected in a decrease in hippocampal activity (Bisby et al., 2016; Murray and Kensinger, 2014; Okada et al., 2011). This negative effect of amygdala activity on hippocampal-dependent association-memory formation is also consistent with a dual-representation account: Better item-memory and worse associative memory for emotional information may be driven by opposing effects of arousal on amygdala- and hippocampal-dependent memory systems (Yonelinas and Ritchey, 2015). Opposing effects of emotional arousal on amygdala and hippocampus, in particular the hypothesized decrease in hippocampal activity, have not yet been specified (Bisby et al., 2016), although likely indirect (via inhibitory/excitatory connections between prefrontal cortex and amygdala versus hippocampus, respectively; Tejeda and

O'Donnell, 2014; Kim et al., 2011; Lee et al., 2012 Moreno et al., 2016). Thus, according to the disruption-hypothesis, the mechanism underlying the memory disadvantage for negative pairs is an indirect disruption of hippocampal associative encoding by emotional arousal.

Alternatively, the 'bypassing-hypothesis,' is based on the observation that when associations can be unitized, association-memory can be supported by extra-hippocampal MTL areas (Haskins et al., 2008; Quamme et al., 2007). Unitization describes the phenomenon that inter-item associations can be merged under certain conditions to function like *intra*-item associations or even processed like a single item. Under these circumstances, their encoding becomes hippocampus-independent and their recognition can be based solely on familiarity (not episodic recollection; Diana et al., 2008; Ford et al., 2010; Giovanello et al., 2006). Unitization seems to be a continuous and not an all-or-none process: The degree of unitization depends on characteristics of the to-be-merged items and the encoding task. For example, it is easier to unitize the color of a word with the word itself than to unitize two sequentially presented same-modality items. Similarly, encoding instructions asking for integrative imagery trigger active unitization attempts more so than non-integrative encoding instructions. Importantly, it has been shown that two neutral items can be encoded without requiring active unitization attempts or instruction, for example, if their combination is by itself meaningful or familiar (Ahmad and Hockley, 2014). Also, if unrelated items belong to the same domain (e.g., face-face pairs) associative encoding can circumvent hippocampal involvement (Bastin et al., 2010; Mayes et al., 2007; Mayes et al., 2004; Tibon et al., 2014). Based on this literature, one could hypothesize that inherently distracting features of emotional items may make them harder to unitize or prevent extra-hippocampal within-domain associations which then might lead to worse association-memory (see also: Mather and Sutherland, 2011; Murray and Kensinger, 2013). Accordingly,

extra-hippocampal MTL activity may be associated with successful neutral but not with successful negative pair encoding The bypassing-hypothesis proposes that the mechanism underlying the memory *advantage* for neutral pairs is additional, extra-hippocampal associative encoding.

Focusing on the amygdala, hippocampus, and extra-hippocampal MTL, different pattern of results can be predicted according to the two hypotheses. To test the prediction of both hypotheses, we examined mean activity during emotional and neutral pair encoding irrespective of subsequent memory as well as subsequent memory effects (SMEs), contrasting brain activity during encoding of later-remembered (hits) vs. later-forgotten (misses) pairs, separately for negative and neutral pairs. Both hypotheses converge with respect to predicting a main effect of emotion in the amygdala: increased amygdala activity during negative than neutral pair encoding. In addition, both hypotheses also predict a subsequent forgetting effect (greater activity during subsequently forgotten than remembered pairs) specifically for the negative pairs; this effect could either be in other parts of the amygdala and/or in stronger coupling between amygdala activity and other brain regions during subsequently forgotten than remembered negative pairs. Thus, using psychophysiological interaction analyses, we also tested potential changes in functional coupling between the amygdala and other brain regions pertaining to forgetting of negative pairs. The disruption hypothesis would predict then together with higher amygdala activity decreased mean hippocampal activity levels during negative than neutral pair encoding. However, this hypothesis would not assume differences in the size of the hippocampal SMEs: Associative encoding is thought to remain hippocampal-dependent and hippocampal activity is equally important to subsequent memory-outcome for negative and neutral pairs, just less likely to occur for the former. Conversely, the bypassing hypothesis assumes given the

higher amygdala activity during encoding of negative compared to neutral pairs no difference in mean-activity levels in the hippocampus. However, because neutral pairs are easier to unitize and amenable to an alternative, extra-hippocampal strategy, this hypothesis predicts that there should be additional SMEs in extra-hippocampal MTL, i.e. the MTL cortex, for neutral pairs that are absent (or weaker) for negative pairs. On an exploratory basis, it might be hypothesized moreover a decrease in mean MTL-cortex activity as a consequence of emotional arousal during encoding of negative arousing pairs is observed.

Following our behavioural paradigm (Madan et al., 2012), we used intentional instructions to maximise the potential of association memories to emerge (Hockley and Cristi, 1996). Experiments 1 and 2 confirmed emotional impairment of association-memory alongside item-memory enhancement (Experiment 2), using a modified procedure of Madan et al. (2012). As our predictions included different response profiles in putatively adjacent MTL regions—amygdala, hippocampus, and MTL-cortex—we scanned the MTL using high-resolution fMRI in Experiment 3. This experiment tests the disruption and bypassing hypotheses with respect to the predicted roles of the MTL regions during encoding of emotional versus neutral associations.

2. Materials and Methods

The study was approved by the local ethics committee, Board of Physicians, Hamburg, Germany. All participants gave written informed consent for this study and received monetary reimbursement (10 €/h). Figure 1 gives an overview of the common features of all three experiments.

2.1. Experiment 1: Adaptation of Madan et al.'s (2012) procedure for fMRI

Several extensive changes were necessary to adapt the original task (Exp. 1 of Madan et al., 2012) for fMRI. Briefly, the original procedure was a verbal paired-associates task, presenting arousing negative and non-arousing neutral words in all possible pairings (pure negative, pure neutral, and mixed pairs). Participants had been explicitly instructed to learn these as pairs and were tested with cued-recall after each of 8 sets of 8 pairs. This was followed by a final free-recall test of all words. Adapting this paradigm for fMRI, we used emotional pictures instead of words, known to elicit more reliable BOLD responses (Kensinger and Schacter, 2006b). Furthermore, the two stimuli of a pair were presented simultaneously to avoid problems with deconvolution of BOLD responses to individual pictures within each pair in the later fMRI task, and to allow meaningful saccadic eye-tracking recordings. To emulate cued recall but avoid vocal recordings in the scanner, participants were first asked to covertly recall the associate of the single probe picture and to make a judgment-of-memory (JoM) with a 2-AFC button-press. This was followed by 5-alternative-forced-choice (5-AFC) associative recognition.

2.1.1. Participants. A total of 42 healthy male volunteers participated in Experiment 1. Participants were right-handed, had normal or corrected-to-normal vision, and reported no past or present psychiatric or neurological disorders. Considering the planned fMRI study (Experiment 3), we selected only males to avoid possible gender-specific lateralization of amygdala activations in tasks involving emotional materials (e.g., Cahill et al., 2004). Data from 6 participants had to be excluded due to below-chance accuracy in the 5-AFC associative recognition task. The final group contained 36 participants.

2.1.2. Experimental Design. A total of 320 pictures (160 negative, 160 neutral) were selected from the International Affective Picture System (Lang et al., 2008) and from the internet. An independent group of 20 male raters from an unrelated study judged arousal-levels of each picture on 9-point modified versions of the Self-Assessment-Manikin scales (Bradley and Lang, 1994). With '9' indicating low arousal, pictures preselected as negative (N) were rated higher in arousal ($M \pm SD = 5.09 \pm 0.85$) than neutral (n) pictures ($M = 7.70 \pm 0.35$; $t(212) = 35.74$, $p < .001$). The experiment was implemented with Presentation (Neurobehavioral Systems Inc.; Berkeley, CA) software.

Experiment 1 comprised three cycles, each with a study phase (Fig. 1A) followed by a test phase (Fig. 1B). Participants first performed five practice trials, with repeats if needed. Excluding the practice pictures, a total of 288 pictures (144 negative, 144 neutral) were randomly selected from the picture pool and presented in three 48-pair cycles.

Insert Figure 1 here

In each encoding trial (Fig. 1A), two pictures (450×300 pixels) were shown side-by-side on a computer screen for 2000 ms (screen resolution 1440×900 pixels), preceded by a fixation cross for 1000 ms. Pictures were shown simultaneously and pairs included all possible permutations of negative (N) and neutral (n) pictures on the left side or the right side of a pair (NN, Nn, nN, nn), as in (Madan et al., 2012), with 12 pairs of each type comprising the 48-pair cycles. Participants were explicitly asked to study the pairings and informed that their memory for each pair would be tested later.

In the retrieval phase, each pair was tested with a JoM task and a 5-AFC associative recognition task (see Fig. 1B). One trial in the JoM task lasted 4900 ms, followed by a 100-ms blank screen and 1000-ms fixation-cross. In the JoM task, pseudorandomized either the left or right picture of the pair, with no more than two repeats of picture emotion, was presented in the center of the screen. Participants were prompted by the question: “Recall associate?” and had to choose a “Yes” or “No” on-screen button with a computer mouse. Participants were asked to be conservative with their memory judgments and to only endorse a ‘yes’ response if they were sure they had remembered the previously associated picture of the pair. For the 5-AFC associative recognition task, the same probe picture was presented in the center of the screen (225 × 150 pixels), surrounded by an array of five pictures (one correct target, four lures) in fixed screen positions (Fig. 1B). Participants had 3900 ms to choose the target picture from the array with a computer mouse, followed by a 100-ms blank screen. Lure pictures were always from the just preceding study phase. The four lures were pseudorandomly selected such that all five recognition alternatives always had a ratio of 2:3 or 3:2 negative to neutral pictures.

An active baseline task was included (Fig. 1C), considering the planned fMRI experiment (Experiment 3), to prevent high resting state brain activity in regions like the hippocampus and therefore avoid possible contamination by task-related activity changes in these regions (Stark and Squire, 2001). Each baseline trial lasted 2000 ms (1900 ms of baseline and 100 ms blank screen). In each baseline trial, a line drawing of a star was presented in one of five screen locations (Fig. 1C), analogous to the picture positions in the 5-AFC task (Fig. 1B). Participants had to select the screen location of the star with the mouse. Two baseline trials were presented after each study trial in the encoding phase and after each associative recognition trial in the retrieval phase. In addition to its function as an active baseline task, this procedure also served as

a test of the participants' ability to accurately choose between the five screen positions as required in the 5-AFC task.

Prior to each encoding phase and retrieval phase, a pictorial two-back task was used to clear working memory and to help participants discriminate between different cognitive contexts (e.g., to separate pictures from the current encoding phase from pictures in earlier encoding phases; Pastotter et al., 2011). The two-back task consisted of 30 trials and lasted 1 minute. The task used five line drawings from Rossion and Pourtois (2004), which were presented sequentially in random order for 1900 ms each, followed by 100 ms of blank screen. Participants were asked to indicate by button press whether the current drawing was a match or no match to the drawing shown two trials prior. Figures 1D and 1E give an overview on the timing of events within the encoding and retrieval phases.

2.2. Experiment 2: Concurrent decrease in association-memory and increase in item-memory for negative pictures

In the substantially modified version of the task, Experiment 1 replicated the basic finding of Madan et al. (2012): an association-memory disadvantage for negative compared to neutral materials (see Results). Item-memory *enhancement* for emotionally arousing information has been well-established, including in many fMRI studies (cf. Dolcos et al., 2012). Our previous study had also identified emotional item-memory enhancement in final free recall (Madan et al., 2012). The goal of Experiment 2 was to test whether these materials and procedure would also produce a simultaneous *increase* in a subsequent item-memory test for individual negative pictures despite a *decrease* in association-memory for negative pairs, similar to our previous findings (Madan et al., 2012). This required the introduction of an item-memory task in the

current design without compromising the intentional associative encoding instruction. The possibility of applying free recall was complicated by the fact that some of the pictures were not uniquely describable. Thus, Experiment 2 contained only one study-test block of pictures, followed by an unannounced 2-alternative-forced-choice (2-AFC) item-recognition memory task. The 2-AFC task presented a previously encoded picture alongside a new lure picture and hence did not require associative encoding/retrieval. This design allowed directly contrasting effects of emotion on association-memory (JoM/5-AFC) with those on item-memory (2-AFC). Contrary to Experiment 1 which aimed to replicate the findings of Madan et al., (2012), in Experiment 2 and 3 only pure neutral and negative pairs were employed to gain statistical power for the comparisons of main theoretical interest. A reduction of conditions was even more important for the experiments that had fewer possible trials (Experiment 2) or where brain activity was measured (Experiment 3). Moreover, pure pairs were expected to reduce differential allocation of attention within a pair.

2.2.1. Participants. A total of 34 healthy male volunteers participated in Experiment 2; six participants were excluded due to below-chance performance in the item-recognition task, retaining 28 participants.

2.2.2. Experimental Design. Of the original 320 pictures from the picture pool, 280 (140 negative and 140 neutral) were selected at random for each participant. Of these, 140 (70 negative/70 neutral) were studied during the encoding phase. A higher number of pictures, compared to encoding blocks in Experiment 1, was necessary to avoid ceiling effects in the 2-AFC. The remaining 140 pictures were used as lure pictures in the 2-AFC item-memory test.

Instead of three encoding-retrieval cycles as in Experiment 1, all 70 pairs were presented in a single cycle. We presented only pure negative (NN) and pure neutral (nn) pairs in Experiment 2, with 35 pairs being presented of each type. Asymmetries in recall from mixed pairs in (Madan et al., 2012) had been attributed to effects of *item*-memory enhancement for negative target words. Similar asymmetries were detected in Experiment 1 here, using mixed pairs. To reduce the number of experimental conditions, we presented only pure pairs in Experiment 2. Since only pure pairs were used, the 5-AFC associative recognition task presented all lures of the same valence (i.e., the alternatives were five negative pictures or five neutral pictures).

The encoding phase, JoM, and 5-AFC associative recognition task were identical to Experiment 1. Participants were again instructed to intentionally encode the pairs. To probe item-memory, an unannounced 2-AFC recognition task was included where all items were tested, preceding the 5-AFC associative-recognition task for all pairs. The 2-AFC task had 140 trials in which a studied, old picture and a non-studied, new lure picture were presented side-by-side for 2900 ms, followed by a blank screen for 100 ms. The new picture was always of the same emotional valence as the accompanying old picture. Participants were instructed to select the studied (old) picture of the two with the computer mouse. The two-back task both preceded and followed the 2-AFC item-recognition task.

2.3. Experiment 3: High-resolution fMRI in medial temporal lobe and eye-tracking during study of negative and neutral pairs

Experiments 1 and 2 replicated an association-memory reduction for negative information and simultaneous item-memory enhancement (Madan et al., 2012). Experiment 3 proceeded to test neural mechanisms underlying both successful and unsuccessful association-memory for

345 negative compared to neutral picture pairs. High-resolution fMRI of the MTL/fusiform regions
346 was used, concentrating on SMEs, i.e., brain activity during encoding of later successfully
347 recognized picture pairs (hits) compared to brain activity during encoding of later-forgotten pairs
348 (misses). In addition, eye-tracking recordings were acquired during encoding to test the potential
349 link between visual attention patterns and later associative memory success/failure. As
350 impairment of association-memory for emotional items might be driven by attentional factors,
351 eye-movements were used as a measure to approximate overt attention.
352
353 **2.3.1. Participants.** A total of 23 healthy right-handed male volunteers participated in
354 experiment 3. Data from 3 participants were excluded due to below-chance performance in the
355 associative recognition task, leaving 20 participants.
356
357 **2.3.2. Experimental Design.** A set of 300 pictures was randomly selected from the original 320
358 pictures for each participant. Similar to Experiment 1, three encoding-retrieval cycles were
359 carried out. These contained 50 pairs in each cycle (25 of each pair type), with a total of 150
360 pairs. As in Experiment 2, only pure negative (NN) and pure neutral (nn) pairs were used and all
361 lure pictures were of the same valence as the target. All other task parameters were identical to
362 Experiment 1. There was no item-memory task.
363 Eye movements were recorded, using a EyeLink 1000 video-based eye-tracker (SR
364 Research Ltd.; Mississauga, ON, Canada), at a sampling rate of 1000 Hz and with a spatial
365 resolution of less than 0.01° and a spatial accuracy of 0.25°-0.4°. An infrared camera located at
366 the edge of the MRI bed was used to monitor participants' eye movements. Eye-tracking data
367 were acquired during encoding and retrieval phases, but only encoding data are presented here.

Six participants could not be included in the eye-tracking analyses due to issues with the eye-tracker reliably detecting their pupils during data collection, leaving 14 participants for the eye-tracking analyses.

Pictures were back-projected onto a screen and viewed through a mirror. Instead of a computer mouse, participants used an MR-compatible joystick (Mag Design and Engineering; Sunnyvale, CA). MR scanning was conducted during both encoding and retrieval phases, but only encoding-related brain activity is presented here. To approximate encoding and retrieval length inside the scanner, the retrieval phase within each cycle was split such that a random set of 25 pairs out of the 50 pairs from the encoding phase was tested in a first retrieval-phase (12-13 neutral or negative pairs), followed by a second retrieval-phase probing memory for the remaining 25 pairs. Thus, 9 experimental runs were conducted in total: encoding (50 pairs), retrieval 1 (25 pairs), retrieval 2 (25 pairs), repeated three times.

2.3.3. MRI data acquisition and analysis. Functional MRI was performed on a 3 T system (Siemens Trio) with an echo-planar imaging T2*-sensitive sequence in 36 contiguous axial slices (1.5-mm isotropic voxels; TR = 2760 ms; TE = 30 ms; flip angle = 80°; field of view = 240×240 mm^2). The field of view was aligned to the longitudinal axis of the hippocampus and covered the temporal lobes as well as part of the insular cortex. Figure 3A illustrates the areas covered by the high-resolution fMRI-sequence. The first five volumes of each functional MR scan were discarded to allow tissue steady-state magnetization. High-resolution T1-weighted structural MR image was acquired by using a 3D-MPRAGE sequence (TR = 2300 ms; TE = 2.89 ms; flip angle = 9°; 1-mm slices; FOV = 256×192; 240 slices).

The functional image time-series was slice-time corrected, realigned and corrected for the interaction of motion and distortion using the unwarp function as implemented in SPM12 (http://www.fil.ion.ucl.ac.uk/spm) which corrects the data for movement related signal changes. Therefore movement regressors were not included in the first level models. Then, the individual structural T1 image was co-registered to the mean functional image generated during realignment using an affine rigid-body transformation and the quality of the co-registration was manually checked for each participant. Co-registered T1 images were segmented using the 'Segment' routine in SPM12. During this step, tissue-class images for gray and white matter were generated from the structural images and subsequently used with the DARTEL toolbox to create individual-subject flow fields, which in turn were used for normalization to MNI space. Functional images were normalized to MNI space using the DARTEL-generated flow fields, re-sliced with an isotropic voxel size of 1 mm, and smoothed with a Gaussian kernel of 3-mm full-width at half-maximum (FWHM) .

Two sets of analyses were conducted. First, we aimed to identify potential differences in mean activity, focussing on the hippocampus (disruption hypothesis) and MTL-cortex (bypassing hypothesis). These analyses included two regressors of interest: neutral and negative pair encoding. Secondly, we tested four regressors of interest to probe SMEs: activity associated with neutral hits, neutral misses, negative hits, and negative misses pairs (see also (Caplan and Madan, 2016).

2.3.3.1. Mean activity analysis. In detail, this analysis was aimed at identifying potential differences in general activity during processing of neutral and negative pairs as suggested by the disruption hypothesis, i.e., a general decrease in hippocampal activity irrespective of encoding

success during processing of negative stimuli (Bisby et al., 2016). First-level models were constructed for each participant with two regressors modeling the onsets of neutral and negative pairs using the SPM canonical hemodynamic response function. To derive noise regressors from voxels unrelated to the experimental paradigm, subject-specific white matter and cerebrospinal fluid masks were generated based on the segmented T1 images. Principal components explaining at least 1% of the variance were extracted independently for white matter and cerebrospinal fluid. These time series were added as nuisance regressors to the first-level models. The parameter estimates of the two regressors of interest, i.e. activity during processing neutral and negative pairs, were contrasted at the second level with participant as a random factor to test whether mean activity in the hippocampus differed in both conditions. Therefore, for each individual participant the mean activity across all hippocampal voxels in both conditions was computed. In addition, we also calculated voxel-wise statistics to test whether and where peak-activity differences were observed within the hippocampal region of interest. Parallel analyses were conducted focussing on MTL-cortex to probe the bypassing hypothesis. For completeness, we also report mean activity differences between negative and neutral pair encoding in the other regions of interests, i.e. the amygdala and fusiform gyrus.

2.3.3.2. Subsequent memory effect (SME) analysis. Next, we aimed to identify activity differences during processing of neutral and negative pairs that were related to successful versus unsuccessful encoding. Thus, another set of first-level models were constructed for each participant, separating pairs further according to subsequent associative recognition hits versus misses (an SME based on the 5-AFC task). The subjective recall judgments in the JoM task were not considered here due to systematic differences between subjective (JoM) and objective (5-

AFC) association-memory performance (see Results). The resulting four conditions (negative associative recognition hits, negative misses, neutral associative recognition hits, neutral misses) were modeled as separate regressors, again using the canonical hemodynamic response function as implemented in SPM. The same nuisance regressors as in the first set of first-level models were included to explain variance related to unspecific noise. In the second-level analyses, activity related to the pair's emotionality, regardless of later recognition success, was identified by contrasting negative and neutral pairs (main effect of emotion). Successful association-memory formation, regardless of the pair's emotionality, was identified by contrasting hits and misses (main effect of memory; 'subsequent memory effect', SME). The first set of analyses was agnostic to memory outcome, simply asking whether activity (e.g., in the hippocampus), was greater or lower during study of NN versus nn pairs. This set of analyses, incorporating memory outcome, enable us to test whether activity within the regions of interest might relate to memory-encoding success. One might think that the main effect of emotion in this set of analyses yields the same information as the mean activity analysis. However, the SME, by its nature, sorts unequal number of trials into the remembered and forgotten conditions. Because average accuracy differed between negative and neutral pairs, the main effect of emotion in the SME analysis is complicated, being a weighted sum of remembered and forgotten trials— where that weighting differs between conditions. Thus, the main effect of emotion in this set of analyses should be interpreted with caution; the measure of activity, apart from later memory-outcome, during study of NN versus nn pairs is directly addressed in the mean activity analysis. To identify brain regions that separated successful association-memory for negative versus neutral pairs, we contrasted brain activity associated with the SME in negative versus neutral pairs by

458 applying both interaction contrasts (Emotion×Subsequent Memory Effect: SME negative > SME
459 neutral; Emotion×Subsequent Memory Effect: SME neutral > SME negative).
460
461 *2.3.3.3. Psychophysiological interaction (PPI) analysis* . A PPI analysis was conducted, as
462 implemented in SPM12, to assess task-related differences in functional coupling between brain
463 regions (Friston et al., 1997). Foreshadowing our results, we tested whether the amygdala
464 subregion involved in emotional processing (main effect of emotion), was more strongly coupled
465 during failed encoding of negative pairs with either the hippocampus (disruption hypothesis) or
466 with extra-hippocampal MTL regions (bypassing hypothesis). Therefore, the seed region was a
467 left amygdala peak functionally defined at the group-level by contrasting negative vs. neutral
468 trials of the SME analysis (see Table 2 and Figure 3; main effect of emotion, $p < .005$,
469 uncorrected, (-19, -7, -15). (Note that the results are consistent when using the amygdala peak
470 from the main effect analysis (-21 -3 -18), see Results.) The time series, as well as the
471 interaction of the time series with the psychological factor, hits vs. misses during encoding of
472 negative pairs, was extracted after adjusting for effects of no interest (including the session
473 constant and high-pass filter). These two time series were included in the new first-level models
474 as additional regressors, and the parameter estimates of the interaction regressors were used in a
475 second-level analysis with participants as a random factor.
476 We also tested whether the differences in functional coupling of the amygdala with the
477 target region co-varied with performance in the associative recognition task: A stronger negative
478 influence of the amygdala on encoding-related regions leading to reduced association memory
479 for negative pairs.
480

481 *2.3.3.4. Regions of interest. A priori* regions-of-interest (ROIs) were based on the two hypotheses
482 of interest. In particular, the amygdalae were selected based on their critical role in processing
483 emotional arousal and in modulating activity in other brain areas during memory formation
484 (Dolcos et al., 2012; Murty et al., 2010). The amygdala-MTL network has been described so far
485 nearly exclusively for emotional item-memory. Nevertheless, these areas were targeted based on
486 their expected roles in emotional associative memory— although with deviating roles— as
487 suggested by the few studies on this topic (Bisby et al., 2016; Murray and Kensinger, 2014). In
488 addition, the hippocampus was chosen based on its well- established role in associative memory
489 processing (Davachi, 2006; Diana et al., 2007; Eichenbaum et al., 2007) which is proposed to be
490 disrupted during encoding of emotional pairs according to the disruption hypothesis (Bisby et al.,
491 2016). The MTL-cortices have been proposed to be involved in memory in a domain-specific
492 manner, in particular in object memory (perirhinal and lateral entorhinal) versus processing
493 scenic or spatial context memory (parahippocampal and medial entorhinal) (Eichenbaum et al.,
494 2012; Schultz et al., 2015; Staresina and Davachi, 2006). The bypassing hypothesis proposes,
495 based on work on the unitization of associations (Quamme et al., 2007) and on within-domain
496 associations (Mayes et al., 2007), that neutral pair-associative memory can be formed also in
497 extra-hippocampal MTL. Unitized pairs of objects or words have been found to be encoded in
498 the perirhinal cortex (Haskins et al., 2008; Staresina and Davachi, 2010), but the lateral
499 entorhinal cortex should be also involved (Eichenbaum et al., 2012; Schultz et al., 2015). The
500 work on within-domain associations suggests that the convergence area of the processing streams
501 of two items in the MTL should be involved in their associative encoding. For the current scenic
502 stimulus material, this convergence area would be the parahippocampal and medial entorhinal
503 cortex. Taken together, based on previous unitization and within-domain association studies, it

was not straightforward to predict a priori which one of the extrahippocampal MTL cortical regions might be most critical for encoding neutral associations here. Therefore, an ROI comprising all three the MTL-cortices was selected, without further segregation. Finally, two regions, the insula and the fusiform gyrus, were included as additional ROIs that are not directly related to the two opposing hypotheses but have been implicated in emotional processing, respectively encoding. The fusiform gyrus shows not only greater activity during associative than item encoding in particular for pictures but also reliably shows enhanced activity during encoding of emotional than neutral information (Kim, 2011; Murty et al., 2010). The part of the insula included in the scan coverage was selected as an additional ROI because it integrates emotional and cognitive processes, and is involved in interoceptive awareness of emotions and bodily states as well as their goal-directed regulation (Chang et al., 2013).

ROIs were manually traced on a T1 image, averaged across all participants, after normalization to MNI space. Ten ROI masks were traced: bilateral amygdala, bilateral hippocampus, bilateral MTL cortices (perirhinal, entorhinal, parahippocampal), bilateral fusiform gyrus, bilateral insula cortex (as included in the scanned slices). ROIs were either traced based on landmarks used in previously published tracing protocols (amygdala, hippocampus, MTL cortex, fusiform gyrus: Franko et al., 2014; Kim et al., 2000; Pastotter et al., 2011; Pruessner et al., 2000; Pruessner et al., 2002) using ITK-SNAP v 2.4.0 (Yushkevich et al., 2006) or published anatomical masks (insula: Deen et al., 2011). Results of all fMRI analyses were considered significant at $p < .05$, family-wise-error (FWE) corrected for multiple comparisons within the *a priori* anatomical ROIs. For exploratory reasons, we also report clusters present within the entire scan volume at $p < .05$-FWE significance threshold with a minimum cluster size of 20 mm^3.

3. Results

3.1. Experiment 1: Adaptation of Madan et al.'s (2012) procedure for fMRI

We conducted a 2×2×2 repeated-measures ANOVA on the accuracy in the 5-AFC associative recognition task with within-subjects factors pair-type (pure pairs, mixed pairs), target-type (negative, neutral), and test direction (forward, backward). Pair-type differentiates whether the studied pair was a pure pair (nn, NN) or a mixed pair (nN, Nn), target-type differentiates whether the to-be-recognized target picture was negative or neutral, and test direction differentiates whether the pair was tested in the forward or the backward direction. For example, encoding a pair of the type 'nN' shows the neutral picture on the left side on the screen and the negative picture on the right. Forward testing of such a pair would use the left item, 'n', as the memory probe picture and asks for recognition of the right item, 'N', as the target picture; backward testing would show the right 'N' as the probe picture and the left 'n' as the target picture (see Madan et al., 2010, 2012, and Madan, 2014, for additional details). Test direction was included to control for potential biases to one side of the screen, such as (right) visual-field preferences for emotional materials (Natale et al., 1983). Results are shown in Figures 2A and 2B.

Insert Figure 2 here

We observed a significant main effect of pair-type ($F(1,35) = 6.28$, $p = .017$), as well as an interaction of pair-type and target-type ($F(1,35) = 28.55$, $p < .001$). Test direction had no main effect on associative recognition and was not involved in any interactions (all p's > .20). Post-hoc tests on the interaction showed that in pure pairs, negative targets were chosen less

accurately than neutral targets ($t(35) = 4.79$, $p < .001$), extending our previous findings of an emotional impairment of association-memory with pictures and a forced-choice associative recognition test, and replicating Bisby et al. (2016). In mixed pairs, negative targets were chosen more accurately than neutral targets ($t(35) = 3.07$, $p < .001$). In addition, accuracy was worse for the pure pairs with a negative target relative to the mixed pairs with a negative target ($t(35) = 2.61$, $p = .01$) and for mixed pairs with a neutral target than for pure pairs with a neutral target than ($t(35) = 5.86$, $p < .001$). This pattern of results directly replicates our previous findings: memory performance was successively worse the more negative items were contained within a pair, an effect previously linked to associative memory reduction (see Madan et al., 2012). Furthermore, target retrievability was superior when the target was negative versus neutral, implying better memory for negative individual pictures, similar to an effect we previously demonstrated to be caused by negative item-memory advantage.

In the JoM task, participants' 'yes' responses, i.e., confidence in their memory, was analyzed with a simplified repeated-measures ANOVA with trial-type (pure negative, pure neutral, mixed) as a within-subjects factor. The main effect of trial-type was significant ($F(2,70) = 14.65$, $p < .001$). Participants were more confident in their memory for pure neutral pairs ($M \pm SD = 0.61 \pm 0.20$) than pure negative pairs ($M = 0.50 \pm 0.23$), with intermediate memory confidence in mixed pairs ($M = 0.55 \pm 0.22$, Bonferroni-corrected post-hoc *t*-tests: all *p's* < .05). 5-AFC associative recognition accuracy contingent on JoM response is reported in Table 1. Of the two measures, 5-AFC associative recognition is a more objective test of memory. Nonetheless, inclusion of the JoM task makes the retrieval process more similar to cued recall, and likely makes the task more hippocampal dependent than if the recognition test solely was

based on the 5-AFC associative recognition test. Performance in the baseline task was at ceiling (> 99% correct trials; response time: M = 766.69 ± 133.61 ms).

Insert Table 1 here

The results in the 5-AFC task closely resemble the previous cued recall results (Madan et al., 2012), namely, a reduction in association-memory for negative pure pairs compared to neutral pure pairs, with intermediate accuracy for mixed pairs but better performance for negative targets. Differences in associative memory accuracy (cued recall in Madan et al., 2012) for different materials can result not just from influences on the association-memory strength, but from effects on the item-level (see also Madan, 2014; Madan et al., 2010). As outlined in detail in Madan et al. (2012), our previous computational model formally tested whether association memory accuracy for negative compared to neutral information was influenced by item-level parameters ('target retrievability,' 'cue effectiveness') or by the association-memory strength itself. The results showed that a net-reduction in accuracy for negative pairs was due to an imbalance of increased item-memory ('target retrievability' model parameter) with a concomitant, larger, decrease of association-memory strength. Here we nominally replicated our previous results with the current design. Importantly, the association-memory impairment must have been large enough to overcome that advantage for negative target-items to produce a net disadvantage for NN pairs. However, because targets were not explicitly recalled, but rather, target options were provided to the participant (the 5-AFC procedure), it is possible that these item-memory effects are not directly related to target-retrievability effects found previously. Experiment 2 addresses this question directly.

595

596 *3.2. Experiment 2: Concurrent decrease in association-memory and increase in item-memory for*
597 *negative pictures*

598 In the 2-AFC task, item-recognition accuracy was higher for negative pictures (M = 0.92 ± 0.07)
599 than neutral pictures (M = 0.89 ± 0.09; $t(27) = 2.35$, $p = .026$; Fig. 2C). As predicted,
600 performance in the 5-AFC task (Fig. 2D) showed the reverse pattern. Since 'test direction' had
601 no influence on the results of Experiment 1, we conducted a simplified analysis comparing
602 accuracy between negative and neutral pairs, without test direction. Associative recognition was
603 worse for negative (NN) pairs (M = 0.31 ± 0.22) than neutral (nn) pairs (M = 0.38 ± 0.29; $t(27) =$
604 2.75, $p = .01$) (see Fig. 2B) [1]. In the JoM task, memory confidence for negative and neutral pairs
605 was not significantly different ($t(27) = 1.46$, $p = .16$), though confidence for neutral pairs was,
606 nominally, slightly higher than for negative pairs (negative: M = 0.32 ± 0.26; neutral: M = 0.36 ±
607 0.27). 5-AFC associative recognition accuracy contingent on JoM response is reported in Table
608 1. Performance in the baseline task was at ceiling (> 99% correct trials; response time: M =
609 686.98 ± 125.03 ms). Thus, Experiment 2 showed that participants were better at item-
610 recognition of negative pictures and thus confirmed positive effect of arousal on the item
611 memory that was suggested by Experiment 1. At the same time participants were worse at
612 associative recognition for negative picture pairs, compared to neutral pictures or neutral pairs,
613 again forming the results of Experiment 1.

614 We next assessed whether these contrasting memory effects were related to each other.
615 Frequencies of individual pictures from each 5-AFC pair that were previously correctly

[1] Accuracy was relatively unaffected by only including pairs where both of the items were successfully remembered in the item-memory test: Associative recognition was worse for negative (NN) pairs (M = 0.32 ± 0.23) than neutral (nn) pairs (M = 0.39 ± 0.30; $t(27) = 3.09$, $p = .005$).

recognized as items (in the 2-AFC task, i.e.: 0, 1, or 2 pictures) were correlated with later 5-AFC association-memory success (1) or failure (0), using Yule's *Q* as a measure of association, which is appropriate for dichotomous variables (Warrens, 2008). *Q* values range from –1 to +1, and can be interpreted much like Pearson correlation. There was no significant relationship between the two types of memory (negative: 95% CI of Yule's *Q* = (–.32, .22); neutral: *Q* = (-.12, .31); CI was calculated via log-odds transform (Bishop et al., 1975; Hayman and Tulving, 1989). Thus, better item-memory for negative than neutral pictures was not related to reductions in association-memory for negative compared to neutral pairs (Fig. 2E), suggesting two different processes, and replicating the findings of the mathematical model in Madan et al. (2012).

In summary, despite substantial changes to the experimental methods from the original study (Madan et al., 2012), including pictures instead of words, presenting the to-be-associated stimuli simultaneously, changes to timing, number of pairs in the encoding/retrieval phases, use of associative recognition instead of cued recall, and the introduction of the JoM task, we were able to replicate in both experiments the basic finding of interest: Worse associative memory for negative compared to neutral pairs. In Experiment 2, we further confirmed that this decrease was accompanied by increased item-memory for negative pictures compared to neutral pictures. The two effects were not related to each other implying separable influences of emotion on item-memory and association-memory. Experiment 3 interrogated the roles, during encoding, of amygdala subregions, hippocampus and other medial-temporal lobe regions in the emotional-arousal impairment of association-memory.

3.3. Experiment 3: High-resolution fMRI in medial temporal lobe and eye-tracking during study of negative and neutral pairs

3.3.1. Behaviour and eye-tracking. Mean 5-AFC associative recognition accuracy of the 20 participants in the fMRI experiment was 0.55 ± 0.16. Similar to Experiments 1 and 2, associative recognition accuracy was lower for negative (NN) pairs (M = 0.53 ± 0.16) than neutral (nn) pairs (M = 0.59 ± 0.17; $t(19) = 3.23$, $p = .004$) (Fig. 2F), again reflecting a net impairment of association-memory due to emotional arousal. Note that there were similar and sufficient numbers of hit and miss trials within each valence, enabling subsequent memory effect analyses of the fMRI data. In the JoM task, subjective memory confidence for neutral pairs (M = 0.48 ± 0.16) was not significantly different from confidence for negative pairs (M = 0.51 ± 0.18; $t(19) = 0.95$, $p = .35$). 5-AFC associative recognition accuracy contingent on JoM response is reported in Table 1. Performance in the baseline task was at ceiling (98% correct; response time: M = 920.58 ± 129.22 ms).

Although the eye-tracking analyses are underpowered because only 14 participants could be analyzed, we included them here to provide additional information about attentional differences in processing of neutral and negative pairs. We tested effects of emotion (negative pairs, neutral pairs), subsequent memory (hits, misses), and their interaction, on two eye-tracking variables: Mean duration of fixations and the number of saccades between the two pictures of a pair. We reasoned that increased fixations of a stimulus reflects depth of processing which should increase item-memory, whereas increased saccades between pictures may support linking them together and increase association-memory. Fixation durations were slightly, although only on trend level significance, longer for negative than neutral pairs ($F(1,13) = 4.10$, $p = .06$). There was no main effect of memory ($F(1,13) = 1.55$, $p = .24$), nor an interaction between emotion and memory ($F(1,13) = 0.37$, $p = .56$) on fixation durations. However, participants made substantially fewer saccades between negative pictures of a pair than between neutral pictures

662 ($F(1,13) = 34.30$, $p < .001$) (Fig. 2G). We also observed more between-picture saccades during
663 encoding of pairs that were later remembered (i.e., hits vs. misses) — a saccade-based
664 subsequent memory effect ($F(1,13) = 5.37$, $p = .037$). The interaction between emotion and
665 memory on between-picture saccades was not significant ($F(1,13) = 0.004$, $p = .95$). Thus, the
666 eye-tracking patterns hinted at deeper processing of negative than neutral images (i.e., longer
667 fixation duration for negative pictures). Saccadic movements between pictures supported later
668 association memory: There were more between-picture saccades for subsequently remembered
669 pairs (hits vs. misses). Importantly there were also fewer between picture saccades for NN than
670 nn pairs.
671
672 **3.3.2. fMRI results.**
673 *3.3.2.1. Mean activity analysis.* The first analysis tested the prediction of the disruption-
674 hypothesis (Bisby et al., 2016), decrease in hippocampal activity due to emotional arousal.
675 Because a general rather unspecific decrease in hippocampal activity is proposed by this
676 hypothesis activity was in a first step averaged across all voxels in the hippocampal ROI. We
677 observed no evidence for a difference in mean activity in the hippocampal ROIs during
678 processing negative and neutral pairs, neither in the left nor right hippocampus (left: $t(19) = 0.00$,
679 $p = .99$; right: $t(19) = 0.08$, $p = .94$; Fig. 3B). To avoid missing any potential differences in
680 hippocampal subregions, voxel-wise statistics were computed as well, but these also revealed no
681 individual voxels with lower activity for the contrast neutral greater than negative in bilateral
682 hippocampus (all *p*s > .5). Thus, no evidence for the disruption hypothesis was observed. To test
683 the bypassing-hypothesis, we compared mean activity in the bilateral MTL-cortex ROI which
684 was lower during negative than neutral pair processing (left: $t(19) = 6.09$, $p < .0001$; right: t(19)

685 = 3.83, $p < .005$; Fig. 3C) . The voxel-based statistical comparison revealed a significant peak in
686 the left MTL cortex (-17 -37 -17), $Z = 5.44$, $p < .001$, $k_E = 522$; and trend in the right MTL
687 cortex (15, -36, -12), $Z = 3.93$, $p = .061$, $k_E = 175$). For completeness, we also compared mean
688 activity in the fusiform gyrus and amygdala ROIs. In the left fusiform gyrus ROI, mean activity
689 was significantly higher during negative than neutral pair encoding ($t(19) = 2.49$, $p < .05$)
690 whereas the right fusiform showed a trend towards a significant difference ($t(19) = 1.99$, $p =$
691 .06). Bilaterally, amygdala activity was higher during negative than neutral pair encoding (left:
692 $t(19) = 5.59$, $p<.0001$; right: $t(19) = 4.30$, $p<.0001$). The voxel-based statistical comparison
693 revealed a significant peak in the left (-21 -3 -18), $Z = 5.79$, $p > 0.001$, $k_E=552$ and right (24 -1 -
694 19), $Z = 5.90$, $p < 0.001$, $k_E = 451$) amygdala. In sum, activity was greater in the amygdala
695 during negative compared to neutral pair encoding, equal in the hippocampus, relatively
696 decreased in the MTL-cortex and increased in the fusiform gyrus.
697
698 Insert Figure 3 here
699
700 *3.3.2.2. Subsequent memory effect (SME) analysis*. Table 2 summarizes the fMRI findings from
701 the analyses that separately modeled effects of both memory and emotion. We observed a main
702 effect of memory (SME) in the left fusiform cortex and the right amygdala, showing greater
703 activity during successful association-memory encoding than during unsuccessful encoding.
704 Additional trends for a SME main effect within the ROIs included activations in the left
705 amygdala, left hippocampus, and right fusiform cortex.
706
707 Insert Table 2 here

We further observed a pronounced main effect of emotion. Regardless of later association-memory success, increased activity was observed during encoding of negative pairs than neutral pairs in large clusters of the bilateral insula (left insula: Fig. 4A) and bilateral amygdala (left amygdala: Fig. 4D). Note that the latter contained the smaller amygdala regions associated with the memory main effect (SME; see Table 2), confirmed by two conjunction analyses (right amygdala: (22, -2, 21); Z = 3.98, p = .03, k_E = 30; left amygdala: (-17, -8, -14); Z = 3.72, p = .065, k_E = 23). Insula activity was localized more specifically to the dorsal and ventral anterior insula according to the connectivity-based atlas by (Deen et al., 2011). The reverse main effects (memory (misses > hits); emotion (neutral > negative)), did not reveal activations within the ROIs, but additional whole-brain results are listed in Table 2.

Participants with a stronger amygdala main effect to negative pairs also tended to visually fixate on individual negative pictures longer than neutral pictures (r = .51, p = .063) and to make fewer saccades between them (r = -.47, p = .09), although these correlations reached only trend-level significance due to reduced statistical power.

Insert Figure 4 here

Critically, we observed an emotion by memory interaction in various ROIs (see Table 2). Inspecting the interaction, successful encoding of negative pairs versus neutral pairs was associated with increased activity in two left hippocampal areas, one anterior and one posterior (Poppenk et al., 2013), and in bilateral insula. The insula peaks were located in its posterior part according to (Deen et al., 2011). Activity in the left insula and in the anterior left hippocampal

cluster are shown in Figures 5B and 5C, respectively. These effects were driven by an SME for negative rather than a subsequent forgetting effect (SFE) for neutral pairs as the bar plots show.

Insert Figure 5 here

Formal follow-up of these interactions showed that there was significantly more activity for remembered than forgotten negative pairs in the hippocampus (anterior Z = 4.62, p = .005; posterior Z = 4.43, p = 0.12) and a trend in the insula (Z = 3.66, p = .087)), but no such differences for neutral pairs (insula: Z = 2.45, p = .84; anterior hippocampus: Z = 1.30, p =.99*;* posterior hippocampus: Z = 0.77, p =.99; p-values FWE-corrected for multiple comparisons).

In contrast, unsuccessful encoding of negative pairs versus neutral pairs was associated with decreased activity in a ventral region of the left amygdala (see Fig. 4C,E), distinguishable from the more central/dorsal amygdala region observed in the main effect of emotion (Fig. 4D), as well as in left MTL-cortex (Table 2). We then formally tested whether the interaction effect in the ventral amygdala more likely represented an SFE to negative pairs or an SME to neutral pairs. That is, we contrasted activity in the two amygdala localizations that showed the interaction effect (-27, -6, -28) and (-22, -6, -27) (Table 2). These rendered some evidence for significant activation differences between remembered and forgotten negative pairs, but no such differences for neutral pairs (negative: Z = 3.83, p = .046; Z = 3.04, p = .39; neutral: Z = 1.76, p = .99; Z = 2.71; p = .63; p-values FWE-corrected for multiple comparisons). Thus, ventral amygdala activity, at least in one of the two identified regions (-27, -6, -28), more likely represents an SFE for negative pairs than an SME for neutral pairs (Fig. 4E).

753 The same logic applied to the interaction effect in the MTL cortex (Fig. 5E). Probing
754 whether this interaction was driven rather by an SME for neutral or by an SFE for negative pairs
755 revealed no significant effects in either of the pair types. Nevertheless, nominally, the pattern of
756 differences implied more of a neutral SME ($Z = 3.71$, $p = .11$; p-values FWE-corrected for
757 multiple comparisons), whereas the negative SFE was not significant ($Z = 2.06$, $p = .9$). Thus,
758 the significant interaction was more likely driven by an SME for neutral than by an SFE for
759 negative pairs. Interestingly, the MTL-cortex interaction peak (-17, -31, -17) was localized very
760 close to the MTL-cortex peak that showed decreased activity due to negative emotion in the first
761 set of fMRI analyses (-17, -37, -17) (compare Fig. 3C and Fig. 5E).
762 Thus, we observed two spatially separable left amygdala activation foci: (a) a more
763 central location associated with negative picture processing irrespective of later memory, and (b)
764 a more ventral location associated with unsuccessful encoding of negative pairs. In addition, we
765 observed an area in the left MTL-cortex where activity correlated more with successful encoding
766 of neutral than of negative pairs.
767
768 *3.3.2.3. Psychophysiological interaction (PPI) analysis.* To test whether there were differences
769 in functional coupling during the processing of negative pairs related to differences in
770 subsequent memory success, a PPI analysis was conducted using the functionally defined left
771 central/dorsal[2] amygdala peak (-19, -7, -15) (Table 2) as a seed region. The PPI identified an area
772 in ventral amygdala (-28, -5, -29) (Z = 3.40, $p = .046$, small-volume-corrected (SVC) based on a
773 sphere with 5-mm radius around the peak activation of the interaction analyses reported above)

[2] 'Central' and 'ventral' amygdala here refer to peak locations within the amygdala ROI. These terms are not meant to imply we measured activity in the central and ventro-lateral nuclei of the amydgala, which cannot be reliably distinguished with the current MRI parameters.

that exhibited stronger functional coupling with the left central/dorsal amygdala seed during encoding of later-forgotten negative pairs than later-remembered negative pairs (i.e., misses > hits). As can be seen in Figure 4, the identified PPI interaction effect spatially overlapped the left ventral amygdala (-27, -6, -28) peak that had shown significant activation differences between remembered and forgotten negative pairs. (We additionally conducted a parallel PPI analysis using the central/dorsal amygdala peak from the mean activity analysis (two-regressor model) (-21, -3, -18) and similarly found a ventral amygdala cluster (-27, -3, -30) ($Z = 3.24$, $p = .048$).) Central/dorsal amygdala activity (negative picture processing) and ventral amygdala activity (unsuccessful encoding of negative pairs) were further positively correlated ($r = .47$, $p = .036$) across subjects. The functional coupling between central/dorsal and ventral amygdala during unsuccessful negative pair encoding was indeed also stronger in people with larger reductions in association memory for negative compared to neutral pairs, although the correlation was only a trend ($r = .41$, $p = .069$).

4. Discussion

In three experiments, we observed consistently lower association-memory for negative compared to neutral pictures in paired-associate tasks. The magnitude of this reduction was comparable across the current experiments (Experiments 1–3: 8.56%, 6.84%, 6.21%, respectively) and the original verbal design (Madan et al., 2012: 7.73%). In addition, we also observed the well-established emotional item-memory enhancement (Experiments 1 and 2). The disruption-hypothesis, that arousal-induced amygdala activity results in decreased hippocampal activity, presumably via the PFC, was not supported. Results were instead consistent with the bypassing-hypothesis: We observed substantially decreased MTL-cortex activity during processing of

797 negative pairs and a stronger SME for neutral pairs in an adjacent area of left MTL-cortex (Fig.
798 5E). Left hippocampal activity (Fig. 5C) was *increased* during encoding of later successfully
799 remembered negative pairs, a finding that was not predicted by either of the two hypotheses.
800 This finding is compatible only with the bypassing-hypothesis, because the disruption-hypothesis
801 explicitly assumes a decrease of hippocampal activity during emotional association-memory
802 encoding (irrespective of encoding success). Moreover, we were able to dissociate two amygdala
803 clusters with distinct response profiles, one in the central/dorsal amygdala linked to negative
804 picture processing irrespective of associative memory encoding success (Fig. 4D) and the other
805 in the lateral/ventral amygdala showing an SFE for negative pairs (Fig. 4C and 4E). The current
806 results suggest that two parallel mechanisms produce the associative memory advantage for
807 neutral over negative pairs: One in the MTL-cortex that exclusively supports successful encoding
808 of neutral pairs, and one in the hippocampus that exclusively supports encoding of negative
809 pairs. This could imply that during negative pair encoding, association-memory supporting
810 hippocampal contributions can only partly compensate for the absence of MTL-cortical
811 contributions, resulting in a net-decrease in association memory for negative pairs.
812
813 **4.1. Neural substrates of emotional associative memory**
814 There is a relatively sparse and methodologically heterogeneous previous fMRI literature
815 on inter-item emotional associative memory (Bisby et al., 2016; Curcic-Blake et al., 2012;
816 Murray and Kensinger, 2014; Okada et al., 2011). The main advance of the current study is the
817 use of a robust and behaviorally grounded paradigm, with multiple replication across
818 experiments. Asking participants directly to encode the associations was rarely done in this field
819 (Berkers et al., 2016; Okada et al., 2011; Onoda et al., 2009), with none of these studies

investigating subsequent memory effects. The only other study using negative picture-picture pairs (Bisby et al., 2016) aimed to test and found support for the disruption hypothesis, implying that increased amygdala activity may disrupt hippocampal activity during negative association memory formation. However, we observed *more* rather than less hippocampal engagement during successful formation of emotional associative memories, which suggests continued and additional engagement of the hippocampus in this difficult task. Identifying subregions within the amygdala that participated in emotional processes versus those involved in forgetting effects further offers novel evidence for neural substrates underlying inferior emotional association memory.

Bisby et al. (2016) interpreted their results as support for the disruption-hypothesis. Briefly, they reported emotional association memory reductions accompanied by reduced anterior hippocampal activity during encoding of negative pairs. Ventral-lateral left amygdala activity promoted subsequent *item*-memory for negative pictures. Together, these results were suggestive of an amygdala-based disruption to hippocampal associative encoding, concurrent with increases to emotional item memory. Methodological differences between Bisby et al. and our study (Exp. 3) may have driven the differences in findings. Notably, Bisby et al. (2016) reported no amygdala main effect to negative pairs, unlike the robust dorsal/central amygdala main effect here. This could point to differences in the scanning resolution and statistical power between studies, the emotional nature of the materials, and/or the emotional involvement of participants (who encoded pairs incidentally in Bisby et al., 2016). Further, the item-memory effect (showing the amygdala-related SME in Bisby et al., 2016) appears to have been based on successful item-memory, but may have included failed association memory responses. As we further did not test item-memory in Experiment 3, these factors taken together make a direct

843 comparison with the current results difficult. Despite these differences, our results cannot
844 support the conclusion that amygdala activity disrupted hippocampal associative memory
845 functions.
846
847 **4.2. Amygdala**
848 The amygdala played a major role in our findings, pointing to differentiable within-
849 amygdala localizations. Negative pictures were linked to stronger central/dorsal activity
850 irrespective of memory. Failed encoding of negative pairs was related to left ventral amygdala
851 activity. Critically, these two effects were functionally coupled, with stronger coupling during
852 encoding of subsequently forgotten than remembered negative pairs as revealed by the PPI
853 where the strength of this coupling marginally correlated with lower negative association-
854 memory performance. Moreover, across participants, those with a larger ventral amygdala SFE
855 also showed more central/dorsal amygdala activity to negative pairs.
856 According to a recent high-resolution fMRI study that aimed to dissociate amygdala
857 subregions, the central/dorsal amygdala cluster identified in our study maps on the basal and
858 centromedial groups, whereas the ventral cluster in our study maps on the lateral nucleus
859 (Hrybouski et al., 2016). Only the centromedial, and to a lesser extent, the basal groups, but not
860 the lateral nucleus, showed enhanced activity in response to negative pictures in Hrybouski et al.
861 (2016), mirroring the response profiles in our study. Based on this combined anatomical and
862 functional consistency, the central/dorsal cluster in our study might reflect activity of the
863 centromedial group and the ventral cluster maps onto the lateral nucleus. The centromedial group
864 receives direct and indirect (via the lateral and basal amygdala) projections from nearly all brain
865 region, in particular from the sensory and prefrontal/orbitofrontal cortex regions and is the main

output region of the amygdala, in particular it also modulates the lateral amygdala (Sah et al., 2003). The lateral amygdala in turn shows - similar to the basal part - strong bidirectional connectivity with the hippocampus and other MTL regions and modulates prefrontal cortex (PFC) (Sah et al., 2003). Acknowledging that even the current high resolution fMRI sequence cannot reliably distinguish sub-amygdalar nuclei, our findings imply that stronger centromedial amygdala responses to negative pairs triggered lateral amygdala activation which then disturbed association-memory formation (via its known projections to the PFC, modulating MTL activity). Future studies including PFC regions should test these suggestions more directly.

The eye-tracking results complement our interpretations of the activity patterns in the amygdala. Longer fixation durations for negative pictures were trend-correlated with central/dorsal amygdala activity. This might reflect an attentional bias towards individual negative pictures, leading to an emotional item-memory advantage (see Experiment 2; Markovic et al., 2014; Pourtois et al., 2013). In contrast, inter-item saccades— a proxy for the distribution of attention between both pictures— supported associative memory. Fewer such saccades were made during negative- than neutral-pair encoding (Fig. 2G) and participants with more central/dorsal amygdala activity to negative pictures also tended to make fewer saccades between them. Thus, emotional arousal might elicit bottom-up attentional processes (longer fixation duration) interfering with attentional processes (fewer saccades) that serve associative encoding, for example, incidental unitization. However, overt attentional processes engaged in *attempts* to encode a pair appear similar regardless of pair-valence, since we did not observe an interaction between emotion and memory in the eye-tracking results. Although these attentional interpretations appear plausible, the eye-tracking results and trends are limited due to low power.

889 **4.3. MTL cortex and hippocampus**

890 MTL-cortex activity at the border between entorhinal and parahippocampal cortex was
891 decreased during negative pair encoding (Fig. 3C) and an area in close proximity was related to
892 successful encoding of neutral, but not negative pairs (Fig. 5E). These results are predicted by
893 the bypassing hypothesis and consistent with findings of non-hippocampal MTL contributions to
894 formation of neutral association memory. Previous studies have suggested better memory for
895 unitized associations in extra-hippocampal MTL cortex, in particular perirhinal cortex. Using
896 verbal materials (Ford et al., 2010; Giovanello et al., 2006; Haskins et al., 2008; Quamme et al.,
897 2007; Staresina and Davachi, 2010) these studies have also shown that unitization can be
898 triggered by as little as forming a combined sentence or artificial compound word. However,
899 irrespective of unitization instructions, Mayes et al. (2004; 2007) suggested that certain types of
900 associations, namely within-domain associations, can be formed by extra-hippocampal MTL
901 regions. According to this work, items can be associated as soon as their processing streams
902 converge in the MTL. For *between*-domain associations, this can only be accomplished by the
903 hippocampus. For within-domain associations, extra-hippocampal regions would be sufficient.
904 The target regions of convergence here, processing two pictures with scenic content, would be
905 the parahippocampal and entorhinal cortices (Eichenbaum et al., 2012; Schultz et al., 2015).
906 Based on these literatures we suggest that the association-memory advantage for neutral pairs
907 could have been driven by better incidental unitization of neutral than negative scenes or more
908 efficient within-domain associative processes, subserved by parahippocampal/entorhinal cortex
909 regions.

910 In addition to evidence in support of the bypassing-hypothesis, we observed hippocampal
911 activity supporting associative encoding of negative pairs. We propose that when sufficiently

arousing information precludes unitization-based or within-domain associative encoding supported by MTL-cortex regions, an alternative, relational hippocampus-dependent encoding strategy may be engaged. Findings outside the emotional memory literature suggest increased hippocampal involvement during encoding with higher memory demands during retrieval (i.e., recollection vs. familiarity, recall vs. recognition, source memory, memory for contextual details, etc.; Beylin et al., 2001; Eichenbaum et al., 2012; Rugg et al., 2012; Smith et al., 2011). Thus, despite the detrimental influence of emotional arousal on associative encoding, negative (but not neutral) pairs accompanied by additional hippocampal activity during encoding were more likely remembered, suggesting that hippocampal activity is partly compensatory.

4.4. Insula

In addition to the MTL regions we focussed on, memory-relevant activations included those in bilateral insula during negative-pair encoding, and in particular, posterior insula during *successful* encoding of negative pairs. Posterior insula, functionally connected with primary and secondary somatomotor cortices is typically related to physical sensations (e.g., pain; Chang et al., 2013). An fMRI meta-analysis by Uddin et al. (2014) illustrated in addition, that apart from substantial co-activation of insular divisions across many tasks and studies, unique activation of the posterior (but not anterior) insula showed a particular involvement in interoceptive awareness (see Uddin et al., 2014). In the current study, posterior insula activity during successful negative-pair encoding could reflect awareness of one's own emotional response to the negative pictures or regulation thereof (Lane et al., 1997; Pollatos et al., 2007; Tsuchiya and Adolphs, 2007; Zaki et al., 2012). Thus, in the current study, successfully forming association memories between two

negative pictures could have required down-regulation of internal emotional states evoked by the individual pictures.

4.5. Conclusions

Association memory for negative information was consistently impaired. Negative information triggered higher central amygdala activity, which modulated ventral-lateral amygdala regions directly linked to failed negative-pair encoding. Only neutral pair encoding benefited from extra-hippocampal contribution, possibly due to easier unitization of neutral than negative information. Counter to previous suggestions, hippocampal activity was not disrupted during negative-pair learning. Instead (left) hippocampus may provide a compensatory role if extra-hippocampal association memory support is not available, supporting association-memory for negative pairs. This increased hippocampal engagement during negative pair learning may partly offset detrimental association memory influences of the amygdala.

Acknowledgements

We would like to thank Frederike Pohlentz for assistance with data collection. This research was supported by a grant from German Research Foundation (DFG SO 952/6-1) to TS, a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to JBC, and by scholarships/fellowships from the DAAD (German Academic Exchange Service), Natural Sciences and Engineering Research Council (NSERC) of Canada, and Canadian Institutes of Health Research (CIHR) to CRM.

References

Ahmad, F.N., Hockley, W.E., 2014. The role of familiarity in associative recognition of unitized compound word pairs. Q J Exp Psychol (Hove) 67, 2301-2324.

Bastin, C., Van der Linden, M., Schnakers, C., Montaldi, D., Mayes, A.R., 2010. The contribution of familiarity to within-and between-domain associative recognition memory: Use of a modified remember/know procedure. Eur J Cog Psychol 22, 922-943.

Berkers, R.M., Klumpers, F., Fernandez, G., 2016. Medial prefrontal-hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity. Neurobiol Learn Mem 134 Pt A, 44-54.

Beylin, A.V., Gandhi, C.C., Wood, G.E., Talk, A.C., Matzel, L.D., Shors, T.J., 2001. The role of the hippocampus in trace conditioning: temporal discontinuity or task difficulty? Neurobiol Learn Mem 76, 447-461.

Bisby, J.A., Burgess, N., 2014. Negative affect impairs associative memory but not item memory. Learn Mem 21, 21-27.

Bisby, J.A., Horner, A.J., Horlyck, L.D., Burgess, N., 2016. Opposing effects of negative emotion on amygdalar and hippocampal memory for items and associations. Soc Cogn Affect Neurosci 11, 981-990.

Bishop, Y., Fienberg, S.E., Holland, P.W., 1975. Discrete multivariate analysis: Theory and practice. MIT Press, Cambridge, MA.

Bradley, M.M., Greenwald, M.K., Petry, M.C., Lang, P.J., 1992. Remembering pictures: pleasure and arousal in memory. J Exp Psychol Learn Mem Cogn 18, 379-390.

Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. J Behav Ther Exp Psychiatry 25, 49-59.

979 Brown, R., Kulik, J., 1977. Flashbulb memories. Cognition 5, 73-99.
980 Cahill, L., McGaugh, J.L., 1998. Mechanisms of emotional arousal and lasting declarative
981 memory. Trends Neurosci 21, 294-299.
982 Cahill, L., Uncapher, M., Kilpatrick, L., Alkire, M.T., Turner, J., 2004. Sex-related hemispheric
983 lateralization of amygdala function in emotionally influenced memory: an FMRI
984 investigation. Learn Mem 11, 261-266.
985 Caplan, J.B., Madan, C.R., 2016. Word imageability enhances association-memory by increasing
986 hippocampal engagement. J Cogn Neurosci 28, 1522-1538.
987 Chang, L.J., Yarkoni, T., Khaw, M.W., Sanfey, A.G., 2013. Decoding the role of the insula in
988 human cognition: functional parcellation and large-scale reverse inference. Cereb Cortex
989 23, 739-749.
990 Curcic-Blake, B., Swart, M., Aleman, A., 2012. Bidirectional information flow in
991 frontoamygdalar circuits in humans: a dynamic causal modeling study of emotional
992 associative learning. Cereb Cortex 22, 436-445.
993 D'Argembeau, A., Van der Linden, M., 2004. Influence of affective meaning on memory for
994 contextual information. Emotion 4, 173-188.
995 Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr Opin
996 Neurobiol 16, 693-700.
997 Deen, B., Pitskel, N.B., Pelphrey, K.A., 2011. Three systems of insular functional connectivity
998 identified with cluster analysis. Cereb Cortex 21, 1498-1506.
999 Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the
1000 medial temporal lobe: a three-component model. Trends Cogn Sci 11, 379-386.

1001 Diana, R.A., Yonelinas, A.P., Ranganath, C., 2008. The effects of unitization on familiarity-
1002 based source memory: testing a behavioral prediction derived from neuroimaging data. J
1003 Exp Psychol Learn Mem Cogn 34, 730-740.
1004 Dolcos, F., Denkova, E., Dolcos, S., 2012. Neural correlates of emotional memories: a review of
1005 evidence from brain imaging studies. Psychologia 55, 80-111.
1006 Dougal, S., Phelps, E.A., Davachi, L., 2007. The role of medial temporal lobe in item recognition
1007 and source recollection of emotional stimuli. Cogn Affect Behav Neurosci 7, 233-242.
1008 Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., Lipton, P., 2012. Towards a
1009 functional organization of episodic memory in the medial temporal lobe. Neurosci
1010 Biobehav Rev 36, 1597-1608.
1011 Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and
1012 recognition memory. Annu Rev Neurosci 30, 123-152.
1013 Ford, J.H., Verfaellie, M., Giovanello, K.S., 2010. Neural correlates of familiarity-based
1014 associative retrieval. Neuropsychologia 48, 3019-3025.
1015 Franko, E., Insausti, A.M., Artacho-Perula, E., Insausti, R., Chavoix, C., 2014. Identification of
1016 the human medial temporal lobe regions on magnetic resonance images. Hum Brain Mapp
1017 35, 248-256.
1018 Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997.
1019 Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218-
1020 229.
1021 Giovanello, K.S., Keane, M.M., Verfaellie, M., 2006. The contribution of familiarity to
1022 associative memory in amnesia. Neuropsychologia 44, 1859-1865.

1023 Haskins, A.L., Yonelinas, A.P., Quamme, J.R., Ranganath, C., 2008. Perirhinal cortex supports
1024 encoding and familiarity-based recognition of novel associations. Neuron 59, 554-560.
1025 Hayman, C.A., Tulving, E., 1989. Contingent dissociation between recognition and fragment
1026 completion: the method of triangulation. J Exp Psychol Learn Mem Cogn 15, 228-240.
1027 Hockley, W.E., Cristi, C., 1996. Tests of encoding tradeoffs between item and associative
1028 information. Mem Cognit 24, 202-216.
1029 Hrybouski, S., Aghamohammadi-Sereshki, A., Madan, C.R., Shafer, A.T., Baron, C.A., Seres,
1030 P., Beaulieu, C., Olsen, F., Malykhin, N.V., 2016. Amygdala subnuclei response and
1031 connectivity during emotional processing. Neuroimage 133, 98-110.
1032 Kensinger, E.A., Corkin, S., 2003. Memory enhancement for emotional words: are emotional
1033 words more vividly remembered than neutral words? Mem Cognit 31, 1169-1180.
1034 Kensinger, E.A., Schacter, D.L., 2006a. Amygdala activity is associated with the successful
1035 encoding of item, but not source, information for positive and negative stimuli. J Neurosci
1036 26, 2564-2570.
1037 Kensinger, E.A., Schacter, D.L., 2006b. Processing emotional pictures and words: effects of
1038 valence and arousal. Cogn Affect Behav Neurosci 6, 110-126.
1039 Kim, H., 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis
1040 of 74 fMRI studies. Neuroimage 54, 2446-2461.
1041 Kim, J.-J., Crespo-Facorro, B., Andreasen, N.C., O'Leary, D.S., Zhang, B., Harris, G., Magnotta,
1042 V.A., 2000. An MRI-based parcellation method for the temporal lobe. Neuroimage 11,
1043 271-288.

1044 Kim, M.J., Loucks, R.A., Palmer, A.L., Brown, A.C., Solomon, K.M., Marchante, A.N., Whalen,
1045 P.J., 2011. The structural and functional connectivity of the amygdala: from normal
1046 emotion to pathological anxiety. Behav Brain Res 223, 403-410.
1047 Kong, L., Chen, K., Tang, Y., Wu, F., Driesen, N., Womer, F., Fan, G., Ren, L., Jiang, W., Cao,
1048 Y., Blumberg, H.P., Xu, K., Wang, F., 2013. Functional connectivity between the
1049 amygdala and prefrontal cortex in medication-naive individuals with major depressive
1050 disorder. J Psychiatry Neurosci 38, 417-422.
1051 Lane, R.D., Fink, G.R., Chau, P.M., Dolan, R.J., 1997. Neural activation during selective
1052 attention to subjective emotional responses. Neuroreport 8, 3969-3972.
1053 Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2008. International affective picture system (IAPS):
1054 affective ratings of pictures and instruction manual. University of Florida, Gainesville, FL.
1055 Lee, H., Heller, A.S., van Reekum, C.M., Nelson, B., Davidson, R.J., 2012. Amygdala-prefrontal
1056 coupling underlies individual differences in emotion regulation. Neuroimage 62, 1575-
1057 1581.
1058 Madan, C.R., 2014. Manipulability impairs association-memory: revisiting effects of incidental
1059 motor processing on verbal paired-associates. Acta Psychol (Amst) 149, 45-51.
1060 Madan, C.R., Caplan, J.B., Lau, C.S., Fujiwara, E., 2012. Emotional arousal does not enhance
1061 association-memory. . J Mem Lang 66, 695-716.
1062 Madan, C.R., Glaholt, M.G., Caplan, J.B., 2010. The influence of item properties on association-
1063 memory. J Mem Lang 63, 46-63.
1064 Markovic, J., Anderson, A.K., Todd, R.M., 2014. Tuning to the significant: neural and genetic
1065 processes underlying affective enhancement of visual perception and memory. Behav
1066 Brain Res 259, 229-241.

1067 Mather, M., 2007. Emotional arousal and memory binding: An object-based framework. Perspect
1068 Psychol Sci 2, 33-52.
1069 Mather, M., Sutherland, M.R., 2011. Arousal-biased competition in perception and memory.
1070 Perspect Psychol Sci 6, 114-133.
1071 Mayes, A., Montaldi, D., Migo, E., 2007. Associative memory and the medial temporal lobes.
1072 Trends Cogn Sci 11, 126-135.
1073 Mayes, A.R., Holdstock, J.S., Isaac, C.L., Montaldi, D., Grigor, J., Gummer, A., Cariga, P.,
1074 Downes, J.J., Tsivilis, D., Gaffan, D., Gong, Q., Norman, K.A., 2004. Associative
1075 recognition in a patient with selective hippocampal lesions and relatively normal item
1076 recognition. Hippocampus 14, 763-784.
1077 Mickley Steinmetz, K.R., Knight, A.G., Kensinger, E.A., 2016. Neutral details associated with
1078 emotional events are encoded: evidence from a cued recall paradigm. Cogn Emot 30, 1352-
1079 1360.
1080 Moreno, A., Morris, R.G., Canals, S., 2016. Frequency-dependent gating of hippocampal-
1081 neocortical interactions. Cereb Cortex 26, 2105-2114.
1082 Murray, B.D., Kensinger, E.A., 2013. A review of the neural and behavioral consequences for
1083 unitizing emotional and neutral information. Front Behav Neurosci 7, 42.
1084 Murray, B.D., Kensinger, E.A., 2014. The route to an integrative associative memory is
1085 influenced by emotion. PLoS One 9, e82372.
1086 Murty, V.P., Ritchey, M., Adcock, R.A., LaBar, K.S., 2010. fMRI studies of successful
1087 emotional memory encoding: A quantitative meta-analysis. Neuropsychologia 48, 3459-
1088 3469.

1089 Natale, M., Gur, R.E., Gur, R.C., 1983. Hemispheric asymmetries in processing emotional
1090 expressions. Neuropsychologia 21, 555-565.
1091 Okada, G., Okamoto, Y., Kunisato, Y., Aoyama, S., Nishiyama, Y., Yoshimura, S., Onoda, K.,
1092 Toki, S., Yamashita, H., Yamawaki, S., 2011. The effect of negative and positive
1093 emotionality on associative memory: an FMRI study. PLoS One 6, e24862.
1094 Onoda, K., Okamoto, Y., Yamawaki, S., 2009. Neural correlates of associative memory: the
1095 effects of negative emotion. Neurosci Res 64, 50-55.
1096 Pastotter, B., Schicker, S., Niedernhuber, J., Bauml, K.H., 2011. Retrieval during learning
1097 facilitates subsequent memory encoding. J Exp Psychol Learn Mem Cogn 37, 287-297.
1098 Pierce, B.H., Kensinger, E.A., 2011. Effects of emotion on associative recognition: valence and
1099 retention interval matter. Emotion 11, 139-144.
1100 Pollatos, O., Schandry, R., Auer, D.P., Kaufmann, C., 2007. Brain structures mediating
1101 cardiovascular arousal and interoceptive awareness. Brain Res 1141, 178-187.
1102 Poppenk, J., Evensmoen, H.R., Moscovitch, M., Nadel, L., 2013. Long-axis specialization of the
1103 human hippocampus. Trends Cogn Sci 17, 230-240.
1104 Pourtois, G., Schettino, A., Vuilleumier, P., 2013. Brain mechanisms for emotional influences on
1105 perception and attention: what is magic and what is not. Biol Psychol 92, 492-512.
1106 Pruessner, J., Li, L., Serles, W., Pruessner, M., Collins, D., Kabani, N., Lupien, S., Evans, A.,
1107 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-
1108 dimensional analysis software: minimizing the discrepancies between laboratories. Cereb
1109 Cortex 10, 433-442.
1110 Pruessner, J.C., Köhler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., Kabani, N., Collins,
1111 D.L., Evans, A.C., 2002. Volumetry of temporopolar, perirhinal, entorhinal and

parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. Cereb Cortex 12, 1342-1353.

Quamme, J.R., Yonelinas, A.P., Norman, K.A., 2007. Effect of unitization on associative recognition in amnesia. Hippocampus 17, 192-200.

Rimmele, U., Davachi, L., Petrov, R., Dougal, S., Phelps, E.A., 2011. Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. Emotion 11, 553-562.

Rossion, B., Pourtois, G., 2004. Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. Perception 33, 217-236.

Rugg, M.D., Vilberg, K.L., Mattson, J.T., Yu, S.S., Johnson, J.D., Suzuki, M., 2012. Item memory, context memory and the hippocampus: fMRI evidence. Neuropsychologia 50, 3070-3079.

Sah, P., Faber, E.S., Lopez De Armentia, M., Power, J., 2003. The amygdaloid complex: anatomy and physiology. Physiol Rev 83, 803-834.

Schultz, H., Sommer, T., Peters, J., 2015. The role of the human entorhinal cortex in a representational account of memory. Front Hum Neurosci 9, 628.

Smith, C.N., Wixted, J.T., Squire, L.R., 2011. The hippocampus supports both recollection and familiarity when memories are strong. J Neurosci 31, 15693-15702.

Staresina, B.P., Davachi, L., 2006. Differential encoding mechanisms for subsequent associative recognition and free recall. J Neurosci 26, 9162-9172.

Staresina, B.P., Davachi, L., 2010. Object unitization and associative memory formation are supported by distinct brain regions. J Neurosci 30, 9890-9897.

1134 Stark, C.E., Squire, L.R., 2001. When zero is not zero: the problem of ambiguous baseline
1135 conditions in fMRI. Proc Natl Acad Sci U S A 98, 12760-12766.
1136 Tejeda, H.A., O'Donnell, P., 2014. Amygdala inputs to the prefrontal cortex elicit heterosynaptic
1137 suppression of hippocampal inputs. J Neurosci 34, 14365-14374.
1138 Tibon, R., Ben-Zvi, S., Levy, D.A., 2014. Associative recognition processes are modulated by
1139 modality relations. J Cogn Neurosci 26, 1785-1796.
1140 Tsuchiya, N., Adolphs, R., 2007. Emotion and consciousness. Trends Cogn Sci 11, 158-167.
1141 Uddin, L.Q., Kinnison, J., Pessoa, L., Anderson, M.L., 2014. Beyond the tripartite cognition-
1142 emotion-interoception model of the human insular cortex. J Cogn Neurosci 26, 16-27.
1143 Warrens, M.J., 2008. On association coefficients for 2x2 tables and properties that do not depend
1144 on the marginal distributions. Psychometrika 73, 777-789.
1145 Yonelinas, A.P., Ritchey, M., 2015. The slow forgetting of emotional episodic memories: an
1146 emotional binding account. Trends Cogn Sci 19, 259-267.
1147 Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-
1148 guided 3D active contour segmentation of anatomical structures: significantly improved
1149 efficiency and reliability. Neuroimage 31, 1116-1128.
1150 Zaki, J., Davis, J.I., Ochsner, K.N., 2012. Overlapping activity in anterior insula during
1151 interoception and emotional experience. Neuroimage 62, 493-499.
1152

Table 1: 5-AFC associative recognition accuracy (*M*±*SD*) contingent on judgement-of-memory (JoM) response, for all experiments.

Pair Type	JoM=Yes	JoM=No	*t*	*p*
Experiment 1				
Pure Negative (NN)	0.83 ± 0.19	0.47 ± 0.22	7.33	< .001
Pure Neutral (nn)	0.87 ± 0.13	0.44 ± 0.19	10.17	< .001
Mixed	0.80 ± 0.17	0.43 ± 0.18	10.53	< .001
Experiment 2				
Pure Negative (NN)	0.47 ± 0.35	0.24 ± 0.16	3.49	< .01
Pure Neutral (nn)	0.47 ± 0.38	0.29 ± 0.20	3.56	< .01
Experiment 3 (fMRI)				
Pure Negative (NN)	0.72 ± 0.18	0.35 ± 0.16	11.66	< .001
Pure Neutral (nn)	0.83 ± 0.10	0.36 ± 0.16	14.06	< .001

Table 2: Regions of interest and whole-brain ANOVA results for the effects of emotion and memory

Region	Peak coordinates (x, y, z)	Z-statistic	Significance	Voxel extent (at *p* = .005)
ROI, small-volume corrected (*p* < . 05)				
Subsequent Memory Effect (SME: Hits > Misses)				
right amygdala	22, -2, -21	3.86	*p* = .047	23
left fusiform	-39, -18, -28	4.13	*p* = .023	47
left amygdala	-17 -9 -13	3.80	*p* = *.054*	21
left hippocampus	-18 -18 -18	3.90	*p* = *.088*	32
right fusiform	24 -47 -20	3.91	*p* = *.073*	28
Emotion (Negative > Neutral)				
left amygdala	-19, -7, -15	5.41	*p* < .001	489
right amygdala	23, -2, -20	5.52	*p* < .001	362
left insula	-42, -4, -1	5.35	*p* < .001	643
right insula	40, 0, -4	5.27	*p* < .001	246
right insula	39, -13, 6	4.08	*p* = .024	31
right insula	38, 8, -10	3.95	*p* = .037	121
Emotion x Subsequent Memory Effect				
(negative: hits > misses) > (neutral: hits > misses)				
left hippocampus	-24, -16, -15	4.63	*p* = .006	39
left hippocampus	-27, -36, -7	4.47	*p* = .011	45
left insula	-45, -11, -1	4.08	*p* = .021	129
right insula	38, -7, -4	4.06	*p* = .025	22
Emotion x Subsequent Forgetting Effect				
(negative: hits > misses) < (neutral: hits > misses)				
left amygdala	-27, -6, -28	3.95	*p* = .033	20
left amygdala	-22, -6, -27	3.88	*p* = .045	30
left MTL cortex	-17, -31, -17	4.03	*p* = .040	17

Region	Peak coordinates (x, y, z)	Z-statistic	Significance	Voxel extent (at *p* = .005)
Whole-brain (FWE, *p* < .05)				
Subsequent Forgetting Effect (Misses > Hits)				
right temporo-parietal junction	50, -51, 31	4.33	*p* = .004	203
left precuneus	8, -73, 35	5.82	*p* < .001	5465
Emotion (Negative > Neutral)				
left inferior temporal gyrus	-45, -49, -15	inf (t = 10.53)	*p* < .001	439
right inferior temporal gyrus	44, -60, -9	inf (t = 10.2)	*p* < .001	1882
right middle occipital	27, -73, 35	5.76	*p* = .002	2349
right thalamus	45, -17, -1	5.31	*p* = .024	637
right hippocampus	23, -41, -2	5.26	*p* = .031	123
Emotion (Neutral > Negative)				
left precuneus	-16, -61, 19	7.07	*p* < .001	12834
right angular gyrus	41, -66, 42	5.97	*p* = .001	3785
left fusiform	-24, -46, -9	5.91	*p* = .001	1762
left middle occipital gyrus	-33, -84, 36	5.50	*p* = .010	2335
right precuneus	2, -64, 44	5.20	*p* = .040	1302

Figure Captions

Figure 1: Experimental procedure of the encoding tasks and associative recognition tasks used in all three experiments. (A) Encoding task with an example trial of a neutral-neutral (nn) pair. B) Recognition task. (C) Baseline task. (D) Timing of the encoding task. (E) Timing of the recognition task. 5-AFC: Five Alternative-Force-Choice associative recognition task; JOM: Judgement of Memory task. In Experiment 2, a 2-AFC item-recognition task for all items occurred between encoding and the 5-AFC associative recognition task for all pairs.

Figure 2: Behavioral results from Experiments 1–3. (A) Accuracy in the associative recognition task (5-AFC) for all negative (NN) and neutral (nn) pairs in Experiment 1. (B) Associative recognition accuracy from all four conditions in Experiment 1: pure negative (NN), pure neutral (nn), and mixed pairs (nN, Nn). For each pair of bars, the left-hand bar plots the forward probe and the right-hand bar plots the backward probe. Gray bars indicate neutral target pictures, red bars indicate negative target pictures. Observe that accuracy for Nn backward is nearly equivalent to nN forward (these tests both have a neutral probe item and a negative target item). Likewise, accuracy is nearly equivalent for Nn forward and nN backward (these tests both have a negative probe and a neutral target) - in turn, lower than Nn-backward and nN-forward. This is what one expects if there is an emotional enhancement of item-memory dependent on the characteristic of the target. That is, both nN and Nn pairs have the same pair composition: one Negative and one Neutral item; thus, within these pairs, there is evidently an effect of item-memory. If we assume that this emotional enhancement of target-item memory is present as well for pure pairs, then the fact that accuracy for nn > accuracy for NN (regardless of probe direction) suggests that there is an emotional impairment of memory for the association that not only cancels out, but surpasses, in magnitude, the emotional enhancement of item-memory. See Madan et al. (2010, 2012) for more discussion of how to interpret such data plots, as well as examples of mathematical model-fits that support these interpretations. (C) Item recognition accuracy in Experiment 2. (D) Associative recognition accuracy in Experiment 2. (E) Proportion

of pairs from Experiment 2 in which two, one, or none of the individual pictures were recognized in the item recognition task, split by associative recognition hits vs. misses. The lack of difference between association-correct and association-incorrect shows that there was no relationship between item- and association-memory. This argues against the possibility that a strong emotional item is the cause of the disruption of association-memory. (F) Associative recognition accuracy in Experiment 3 (fMRI). (G) Mean number of saccades between the two pictures of a pair in Experiment 3 for remembered (Hit) and forgotten (Miss) negative (NN) and neutral (nn) pairs. Chance level in the 5-AFC associative recognition task was 1/5 = 0.20 (panels A, B, D, F). Chance level in the 2-AFC forced choice item-recognition task was 1/2 = 0.50 (panel C). Error bars are 95% confidence intervals around the mean, corrected for inter-individual differences (Loftus and Masson, 1994).

Figure 3: MRI acquisition and region-of-interest (ROI) results from Experiment 3. (A) Sagittal and coronal sections from the MPRAGE anatomical volume (1 mm^3) illustrating the functional scan coverage in the fMRI study. Mean encoding activity for (B) hippocampal and (C) MTL cortex ROIs, regardless of memory outcome.

Figure 4: Activations and beta estimates from Experiment 3. (B) Coronal slice showing activation clusters. (A) Main effect of emotion in the left insula and (D) left central amygdala. (C,E) Emotion x SME interaction in the left ventral amygdala. Conditions are denoted as negative-negative (NN) or neutral-neutral (nn) pairs that were either hits or misses in the associative recognition task. PPI = psychophysiological interaction analysis with left central/dorsal amygdala seed. Blue region indicates a ventral amygdala region showing significant functional coupling to the seed region, $p = .04$, small-volume-corrected.

Figure 5: Subsequent memory effects (SME) interaction results from Experiment 3. (A) Coronal slice showing the SME clusters specific to negative pairs. Beta estimates are shown for clusters in the (B) left posterior insula and (C) left hippocampus. (D) Coronal slice showing SME clusters specific to neutral pairs. (E) Beta estimates for cluster in the left MTL cortex.

Figure 1

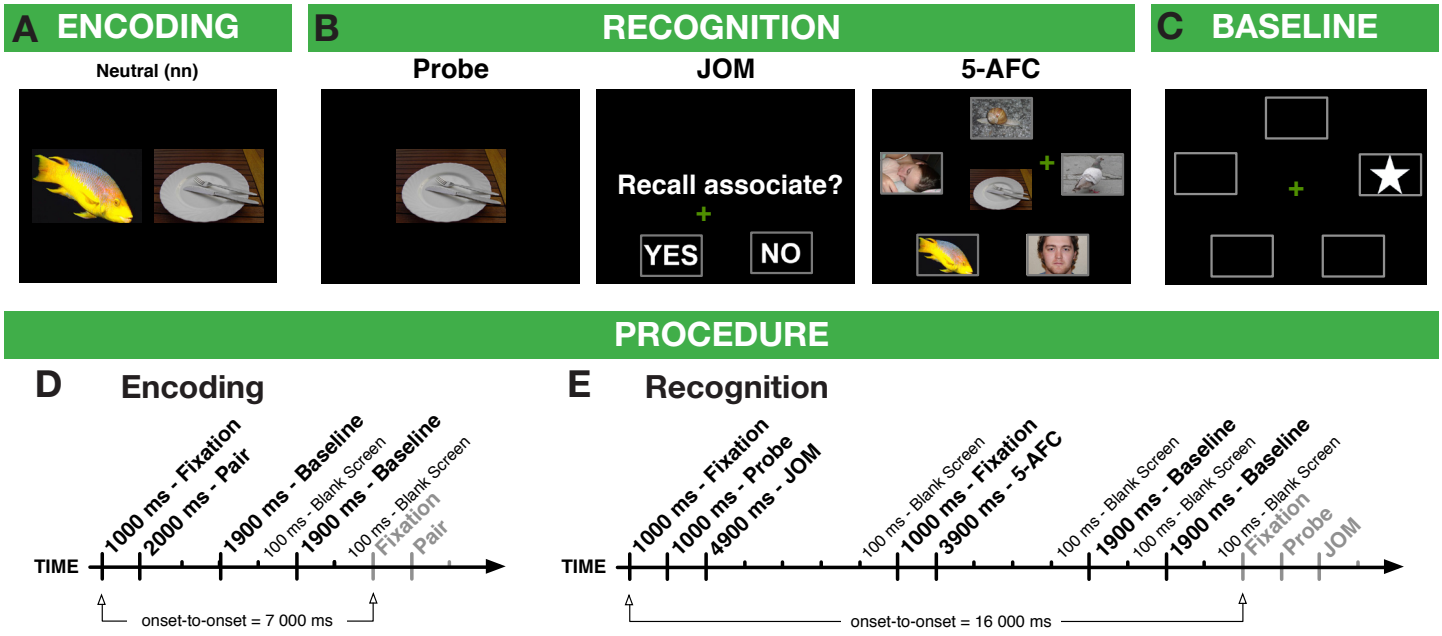

A ENCODING
Neutral (nn)
B RECOGNITION
Probe
JOM
Recall associate?
YES
NO
5-AFC
C BASELINE
PROCEDURE
D Encoding
TIME
1000 ms - Fixation
2000 ms - Pair
1900 ms - Baseline
100 ms - Blank Screen
1900 ms - Baseline
100 ms - Blank Screen
Fixation
Pair
onset-to-onset = 7 000 ms
E Recognition
TIME
1000 ms - Fixation
1000 ms - Probe
4900 ms - JOM
100 ms - Blank Screen
1000 ms - Fixation
3900 ms - 5-AFC
100 ms - Blank Screen
1900 ms - Baseline
100 ms - Blank Screen
1900 ms - Baseline
100 ms - Blank Screen
Fixation
Probe
JOM
onset-to-onset = 16 000 ms

Figure 2

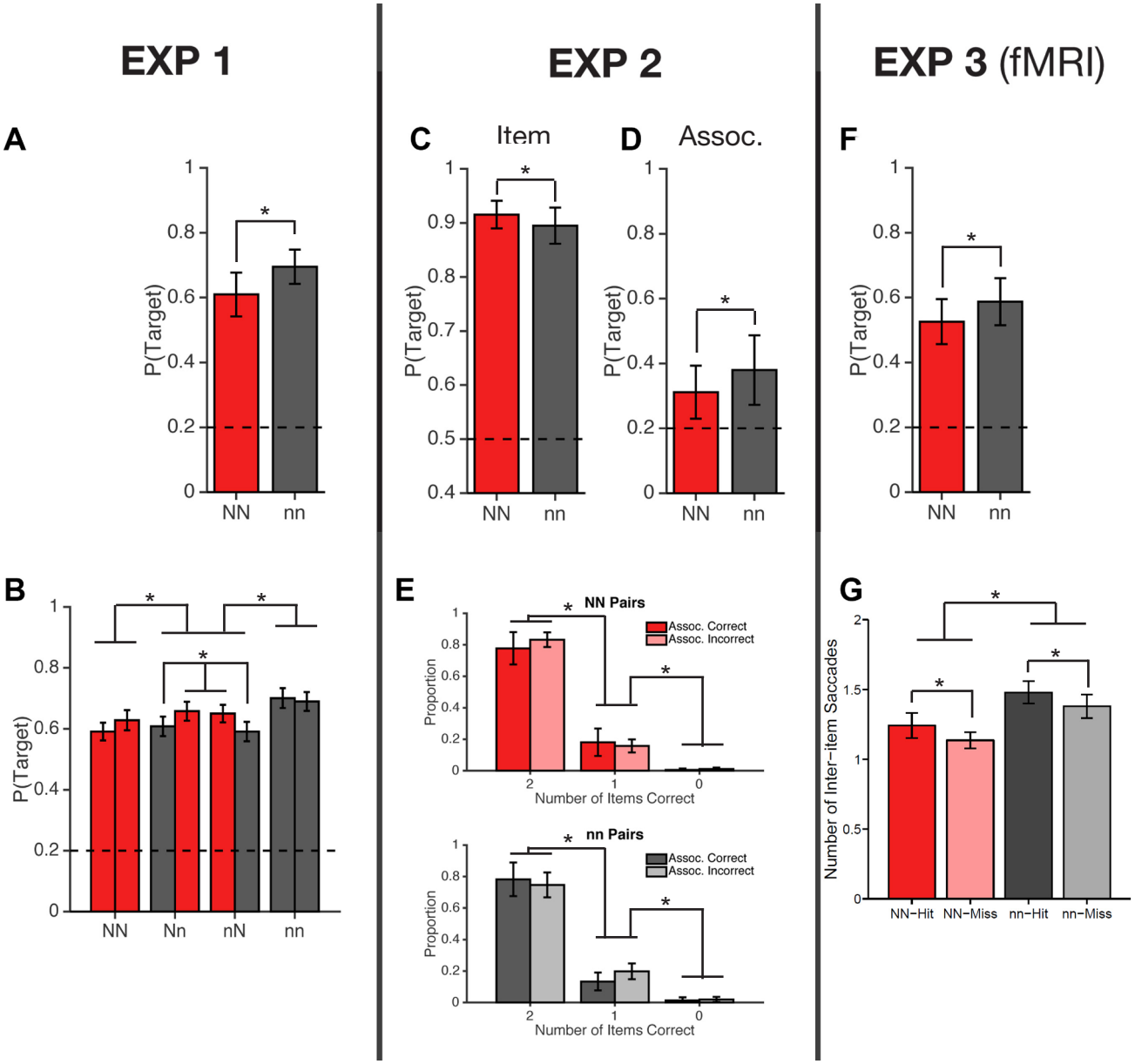

EXP 1
EXP 2
EXP 3 (fMRI)
A
B
C
Item
D
Assoc.
E
F
G
P(Target)
NN
nn
Nn
nN
NN Pairs
nn Pairs
Assoc. Correct
Assoc. Incorrect
Proportion
Number of Items Correct
Number of Inter-item Saccades
NN-Hit
NN-Miss
nn-Hit
nn-Miss

Figure 3

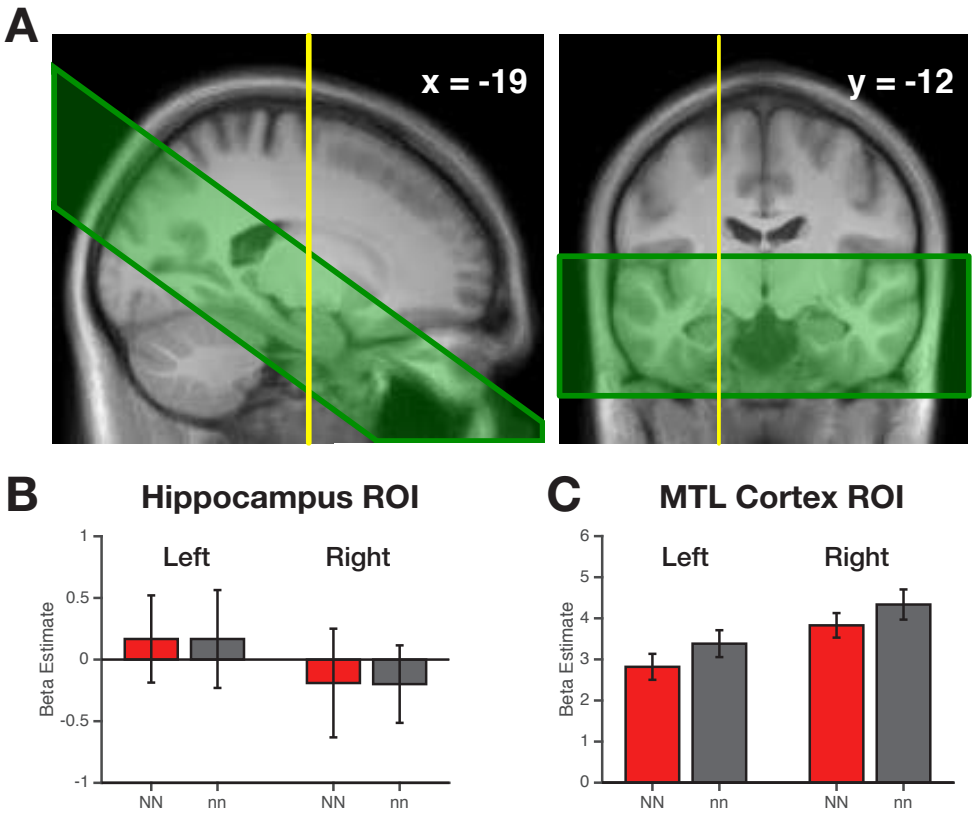

A
x = -19
y = -12
B
Hippocampus ROI
Left
Right
Beta Estimate
1
0.5
0
-0.5
-1
NN
nn
NN
nn
C
MTL Cortex ROI
Left
Right
Beta Estimate
6
5
4
3
2
1
0
NN
nn
NN
nn

Figure 4

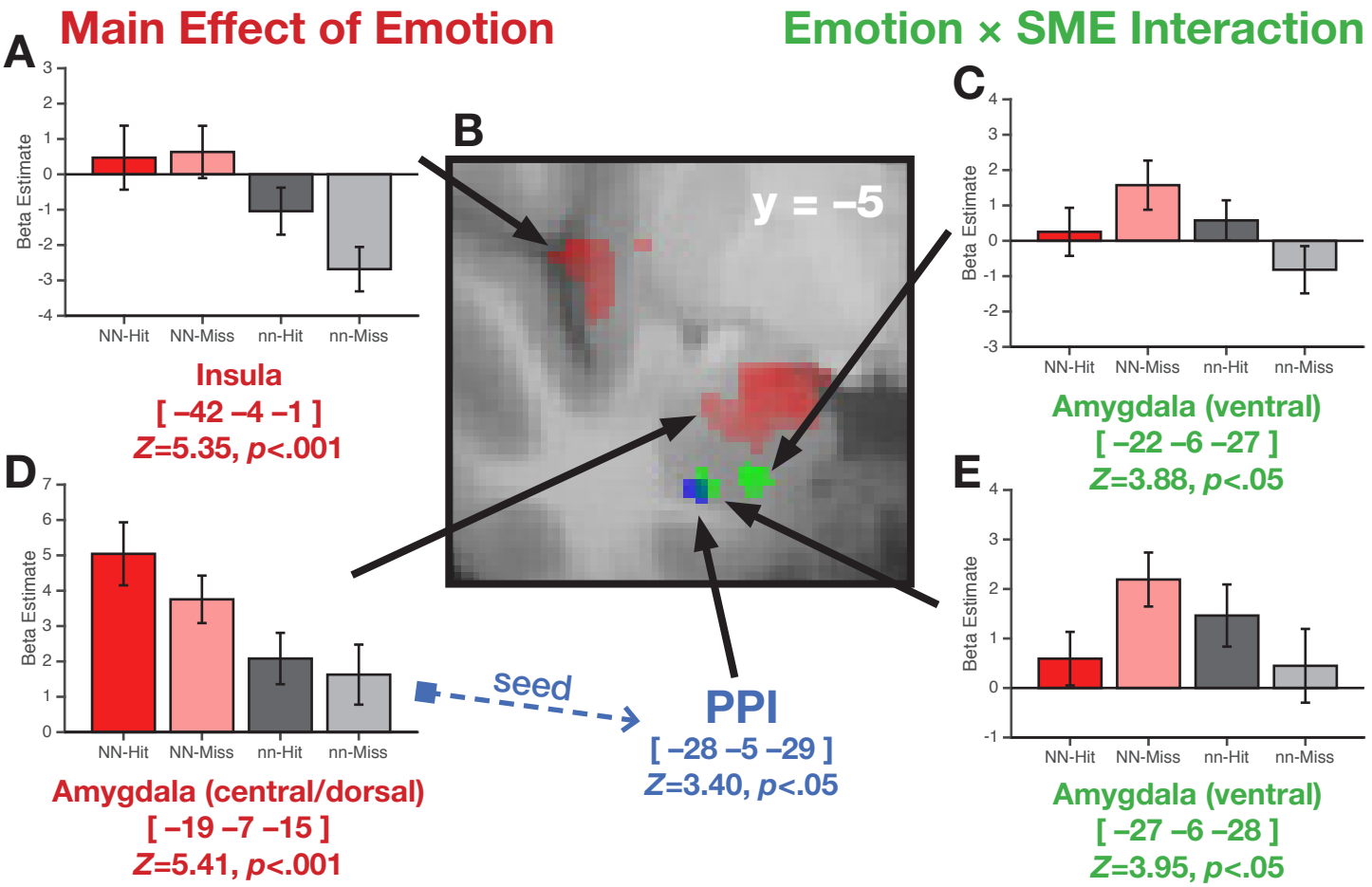
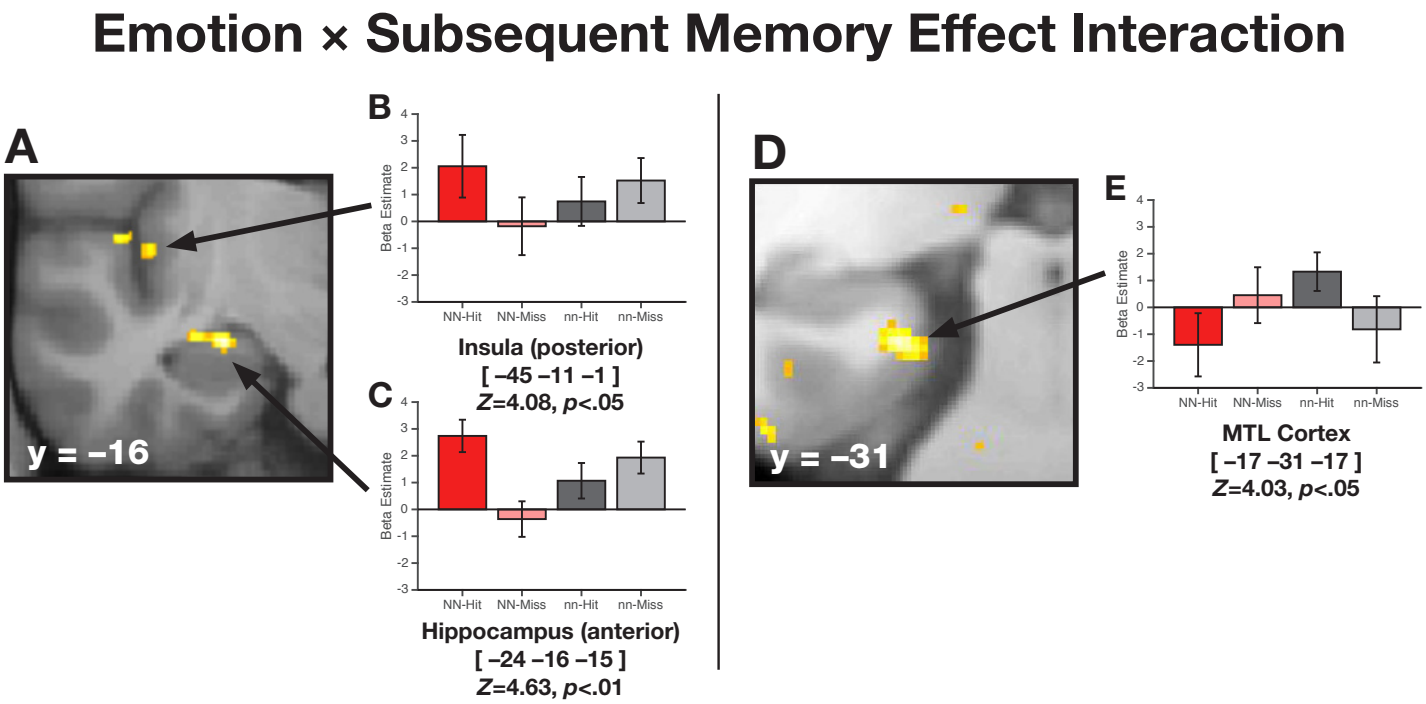

Main Effect of Emotion
Emotion × SME Interaction
A
B
C
D
E
y = –5
Beta Estimate
NN-Hit
NN-Miss
nn-Hit
nn-Miss
Insula
[–42 –4 –1]
Z=5.35, p<.001
Amygdala (ventral)
[–22 –6 –27]
Z=3.88, p<.05
Amygdala (central/dorsal)
[–19 –7 –15]
Z=5.41, p<.001
seed
PPI
[–28 –5 –29]
Z=3.40, p<.05
Amygdala (ventral)
[–27 –6 –28]
Z=3.95, p<.05

Figure 5


Emotion × Subsequent Memory Effect Interaction
A
y = –16
B
Beta Estimate
NN-Hit
NN-Miss
nn-Hit
nn-Miss
Insula (posterior)
[–45 –11 –1]
Z=4.08, p<.05
C
Hippocampus (anterior)
[–24 –16 –15]
Z=4.63, p<.01
D
y = –31
E
MTL Cortex
[–17 –31 –17]
Z=4.03, p<.05