Neural mechanisms underlying subsequent memory for personal beliefs: An fMRI study

Erik A. Wing[1,*], **Vijeth Iyengar**[1,*], **Thomas M. Hess**[2], **Kevin S. LaBar**[1], **Scott A. Huettel**[1], and **Roberto Cabeza**[1]

[1]Center for Cognitive Neuroscience, Duke University, Durham, NC 27708

[2]Department of Psychology, North Carolina State University, Raleigh, NC 27695

Abstract

Many fMRI studies have examined the neural mechanisms supporting emotional memory for stimuli that generate emotion rather automatically (e.g., a picture of a dangerous animal or of appetizing food). However, far fewer studies have examined how memory is influenced by emotion related to social and political issues (e.g., a proposal for large changes in taxation policy), which clearly vary across individuals. In order to investigate the neural substrates of affective and mnemonic processes associated with personal opinions, we employed an fMRI task wherein participants rated the intensity of agreement/disagreement to socio-political belief statements paired with neural face pictures. Following the rating phase, participants performed an associative recognition test in which they distinguished identical vs. recombined face-statement pairs. The study yielded three main findings: behaviorally, the intensity of agreement ratings was linked to greater subjective emotional arousal, as well as enhanced high-confidence subsequent memory. Neurally, statements that elicited strong (vs. weak) agreement or disagreement were associated with greater activation of the amygdala. Finally, a subsequent memory analysis showed that the behavioral memory advantage for statements generating stronger ratings was dependent on the medial prefrontal cortex (mPFC). Together, these results both underscore consistencies in neural systems supporting emotional arousal and suggest a modulation of arousal-related encoding mechanisms when emotion is contingent on referencing personal beliefs.

Introduction

Imagine that while walking on a trail you almost step on a venomous snake. Now imagine that on a different day you hear that your state legislature supports increased public surveillance, which you strongly believe violates personal liberties. These simple examples illustrate the wide range of events that can drive an affective response. Whereas the emotions elicited by the snake may occur in a relatively rapid and automatic fashion, in other cases, emotion results from the appraisal of more abstract personal beliefs. Rather than being shared by most people (e.g., a similar fear response to a snake), belief-related emotions may be highly variable across individuals, as they depend upon referencing one's personal values

Correspondence should be addressed to Erik A. Wing, Center for Cognitive Neuroscience, Duke University, LSRC Bldg, Rm B254, Box 90999, Durham, NC 27708. Phone: (919) 668-2299. eaw31@duke.edu.
*denotes equal contribution

or experiences. For example, a political statement that is outrageous for one person may sound completely innocuous, or even positive, to another person.

Such individual differences make belief-related emotions more difficult to examine, which is one of the main reasons why many studies exploring the neural correlates of emotion instead use stimulus sets (such as the *International Affective Picture System*—IAPS) that include pictures of, for example, dangerous animals or appetizing food. Despite some variability, emotional responses in these studies are more likely to be stimulus-driven and to exhibit some measure of consistency across individuals (Lang, Bradley, & Cuthbert, 1997). By contrast, to our knowledge, only a few studies (e.g., Bruneau & Saxe, 2010; Gozzi, Zamboni, Krueger, & Grafman, 2010; Harris et al., 2009) have investigated brain function in the context of belief-related emotions (e.g., based on differences in ethnic, religious or political dimensions). Thus, while past research on affective processing has helped characterize the function of regions like the amygdala and wider limbic system (Phelps & LeDoux, 2005), much of our current knowledge comes from studies of emotional arousal in the context of lower-level affective stimuli. Although some previous work has explored, for example, the connection between interest in politics and brain responses in the amygdala and striatum (Gozzi, Zamboni, Krueger, & Grafman, 2010), brain processes associated with belief-related emotional arousal remain relatively understudied.

Along with producing substantial arousal, some research suggests that belief-related emotions may produce behavioral memory enhancements (Bradley, Angelini, & Lee, 2007; Civettini & Redlawsk, 2009), which have long been observed in studies using less complex emotional stimuli (Cahill & McGaugh, 1995; Dolcos, LaBar, & Cabeza, 2004; for a review see LaBar & Cabeza, 2006). Such studies typically report better memory for both positive or negative (vs. neutral) stimuli, indicating that the critical factor is emotional arousal, not emotional valence (Hamann, 2001; LaBar & Cabeza, 2006). Corresponding neuroimaging work has shown that this enhancement is associated with increased activity in the amygdala (Dolcos, LaBar, & Cabeza, 2004a; Dolcos et al., 2004b; Dolcos, LaBar, & Cabeza, 2005). In the case of belief-related emotion, however, it is unclear if potential memory-enhancing effects are primarily related to emotional arousal or if emotional valence also plays a role.

In addition to the amygdala, findings from several related lines of fMRI research suggest that the process of evaluating and remembering personally relevant information involves the medial prefrontal cortex (mPFC) (Delgado et al. 2016; Amodio & Frith, 2006; Kelley et al., 2002; Wagner, Haxby, & Heatherton, 2012). For example, social neuroscience work on impression formation has found that various areas within the frontal midline are sensitive to dimensions like the similarity between oneself and others (Leshikar, Cassidy, & Gutchess, 2016; Mitchell, Macrae, & Banaji, 2006) and to the social relevance of stimuli or task orientation (Gilron & Gutchess, 2012; Mitchell, Macrae, & Banaji, 2004). Importantly, these and related studies have also shown that the mPFC influences subsequent memory for social impressions and self-relevant information (Gilron & Gutchess, 2012; Harvey, Fossati, & Lepage, 2007; Macrae, Moran, Heatherton, Banfield, & Kelley, 2004; Mitchell et al., 2004). Subsequent memory effects (SMEs) in mPFC have also been shown for material that is congruent with one's prior knowledge, or pre-existing schemas (Brod, Lindenberger, Wagner, & Shing, 2016; van Kesteren et al., 2013). By contrast, it has been proposed (van

Kesteren, Ruiter, Fernández, & Henson, 2012) that the hippocampus—which plays an important role during episodic encoding more generally—may preferentially support memory for novel information (Kumaran & Maguire, 2009) or arbitrary associations (Konkel & Cohen, 2009). In fact, several studies involving social judgments or self-reference have found that SMEs were characterized by activity in mPFC rather than the hippocampus (Gilron & Gutchess, 2012; Macrae et al., 2004), which in one case preferentially responded to memory during a non-social sequence task on similar stimuli (Mitchell et al., 2004). Taken together, this evidence points toward a particular mnemonic role for the mPFC in contexts where information is relevant along social or schematic dimensions and suggests it may underpin memory in the present study, where pre-existing personal beliefs are evaluated in response to the beliefs of other individuals.

The current study had three main goals. Our *first goal* was to test a behavioral paradigm for investigating belief-related emotions that: **(a)** was both ecologically valid and suitable for fMRI, **(b)** was internally and externally valid, **(c)** elicited significant emotional arousal, and **(d)** enhanced subsequent memory. Whereas previous fMRI paradigms have presented belief statements (e.g. Bruneau & Saxe, 2010; Gozzi et al., 2010), these paradigms did not examine differential effects of belief evaluation on subsequent memory. To address goals A–C, we ran a separate behavioral normative study to identify a set of belief statements that covered a broad spectrum of socio-political issues, yielded desirable levels of internal/external validity, and generated emotional arousal (see Methods and Results). For the scanned task (Figure 1 referred hereafter as the *Agreement with Beliefs Task* —ABT), each belief statement was paired with an expressionless, unfamiliar face while participants rated the extent to which they agreed or disagreed with the statement (four-point scale encompassing strongly or weakly agree or disagree: SA/WA/WD/SD, with separate 'Don't Know' option). A separate associative recognition task phase followed this ABT phase. In the associative recognition task, participants were presented with intact face-belief statement pairs (*identical pairs*) or with pairs in which the face and accompanying belief statement came from different ABT trials (*recombined pairs*). Participants responded whether each pair was identical or recombined, and then made a confidence judgment about their decision. Responses to intact trials were then used to label agreement decisions trials from the ABT phase as either subsequently remembered or subsequently forgotten.

Our *second goal* was to investigate the neural mechanisms of emotional arousal associated with responses to belief statements. After confirming the link between emotional arousal and the intensity of agreement/disagreement responses in a separate behavioral normative study, the corresponding fMRI analyses identified emotional arousal-related activity by comparing high-arousal (SA/SD) versus low-arousal (WA/WD) judgments. Given that a wide range of stimuli are capable of producing affective responses in the amygdala (visual, auditory, olfactory; for a review see Zald, 2003), it is reasonable to expect that emotional arousal elicited by strong agreement or disagreement with socio-political statements would also evoke robust activity in this region. Indeed as mentioned before, political stimuli can be emotional in nature and have been shown to elicit amygdala responses in individuals interested in politics (Gozzi et al., 2010). Nevertheless, it is also possible that amygdala activity would not reflect differences in agreement intensity to short statements attributed to

unfamiliar neutral faces. Although affective responses to a range of environmentally salient stimuli include rapid and sometimes automatic engagement of the amygdala (Diano, Celeghin, Bagnis, & Tamietto, 2017; Phelps & LeDoux, 2005), arousal in the present context is contingent upon parsing relatively abstract information in the context of personal beliefs. This may involve increases in frontally mediated self-referential processing or other operations related to statement evaluation. Therefore, in addition to focusing on the amygdala as a region-of-interest (ROI), we also conducted an exploratory whole-brain analysis to investigate the effects of agreement intensity in other brain regions.

Finally, our third and most important goal was to investigate the neural mechanisms underpinning any enhancement in associative memory formation stemming from belief-related emotions. Using the subsequent memory paradigm (Paller & Wagner, 2002), we identified regions showing greater encoding activity for face-statement pairs associated with higher memory scores during associative recognition, or *subsequent memory effects* (SMEs). In particular, we were interested in how later memory might vary as a function of agreement intensity, which would suggest an influence of belief-related emotional arousal on mnemonic encoding. Given the stimulus and task-related dimensions under which previous work has uncovered medial frontal SMEs, we hypothesized that the mPFC might support memory enhancement in the ABT, where emotional arousal stems from referencing personally held opinions.

Methods

Participants

Twenty-one young adults (11 women; mean age = 21.52 years, SD = 2.40 years) participated in the behavioral normative study and twenty-eight young adults in the fMRI study. fMRI participants were healthy, right-handed, native English speakers, with no disclosed history of neurological or psychiatric episodes. Four fMRI participant were excluded from the analyses due to a high rate of missing or 'Don't Know' encoding responses (>2SD above mean). One participant with motion in excess of the voxel size (3.8mm) was also removed, along with one subject that switched response options during the experiment making responses uninterpretable. The final sample contained twenty-two participants (11 women; mean age = 23.22 years, SD = 3.66 years; years of education = 15.55 years, SD = 1.77 years). All participants in each study gave written informed consent for protocols approved by the Duke University institutional review board.

The number of participants in the behavioral norming and neuroimaging studies was based on effect estimates and sample size from closely related studies on self-referential memory enhancement and mPFC function in social memory paradigms. In the context of behavioral work, the current experiments were adequately powered to detect effects of sizes similar to those reported in two meta-analyses of self-reference memory benefit—($d = 0.63$; Bentley, Greenaway, & Haslam, 2017) and ($d = 0.65$; Symons & Johnson, 1997). Our final fMRI sample ($n = 22$) was also comparable to or greater than n's used in similar neuroimaging studies. These studies—discussed in the Introduction and Discussion—include aspects of both social/self-related processing and episodic memory (mean $n = 19.17$, range $n = 17–22$)

(Gilron & Gutchess, 2012; Harvey et al., 2007; Kelley et al., 2002; Leshikar et al., 2016; Macrae et al., 2004; Mitchell et al., 2004).

Stimuli

Stimuli consisted of 180 face-belief statement pairs. Face pictures of unfamiliar male and female individuals with neutral expressions were obtained from an online database (http://agingmind.utdallas.edu/facedb/view/neutral-faces). All face pictures were presented in color on a black background. The belief statements, which ranged from 4–14 words (M = 9.05, SD = 1.84) and 32–89 characters (M = 58.89, SD = 10.79), consisted of an assortment of viewpoints on a spectrum of social (e.g., abortion, death penalty, immigration, etc.) and economic (e.g., public spending, taxes, etc.) issues. Belief stimuli presented to participants were either modified from stimuli used in a previously published neuroimaging study (Zamboni et al., 2009), adapted from an online test gauging political attitudes (http://www.politicalcompass.org/test), or were novel stimuli that we created in the laboratory for the purposes of this investigation. For every belief statement (90 stimuli), a statement expressing the opposing or counter-viewpoint (90 stimuli) was constructed in order to assess internal consistency and increase trial count. For example, if the original statement read, '*Thinks the law should limit experiments with human embryos*,' its counter-viewpoint would read '*Supports increased funding for stem cell research.*' The original viewpoint and counter-viewpoint statements shared few words in common to reduce memory interference between them.

All belief statements were presented in white text on a black background and appeared beneath face pictures. The pairing of faces and statements was fixed, but was arranged in such a way as to balance demographic features of faces across statements of each political viewpoint (liberal, conservative, independent). To compress variance associated with race, Caucasian faces with neutral expressions were used for all face pictures. As the age of the individuals depicted in the face stimuli was available (face age range: 18–93 years), we distributed face pictures representing different age segments equally among political statement categories, and the same was done for sex, resulting in an equal number of male and female faces across the entire study.

Behavioral methods

Behavioral normative study—The goal of this ancillary study was to gain a better understanding of the emotional dimensions comprising the agreement/disagreement-related judgments. A separate group of participants who did not take part in the imaging study were first presented with the socio-political belief statements used in the scanned-ABT and instructed to provide self-paced agreement responses on a similar 4-pt scale ranging from 'strongly disagree' to 'strongly agree' (or to respond 'don't know'). Instructions on using the 4-pt scale and the 'don't know' response options were identical to those provided to participants in the imaging study. Unlike in the imaging version of the task, after making each agreement rating, participants were further prompted to provide ratings of emotional arousal and issue knowledge, which occurred in a random order for each statement. The emotional arousal rating was on a 4-pt scale ranging from 1-'does not move me,' to 4-'moves me quite a bit' with a 'don't know' response option. Analogous response options

(from 1-'not very knowledgeable', to 4-'very knowledgeable') were used for the knowledge question, which also included a 'don't know' option. Through the study, participants were encouraged to use the entire rating scale while making responses based on their personal political beliefs. Following the study, participants completed a brief written questionnaire. In the questionnaire, they were asked to indicate which of the following terms—*very liberal, somewhat liberal, somewhat conservative, very conservative*—best described their political orientation. For participants in the fMRI study, this question was included in an online survey that participants completed within one month of their scan session.

fMRI study—Prior to entering the scanner room, participants were introduced to the task and took part in an abridged practice version of the scanned paradigm so they would be familiar with task instructions and experimental conditions. The scan session consisted of twelve functional runs alternating between ABT and associative recognition task runs (Figure 1). During each ABT run, participants were presented with 30 unfamiliar face-belief pairs and instructed to rate the degree to which they either disagreed or agreed (1 = strongly disagree, 4 = strongly agree) with the viewpoint ascribed to the accompanying person. Participants were further instructed to choose the 'don't know' response option if they held no knowledge of the issue contained in the statement. These types of responses were subsequently excluded from further analysis. Trials in ABT runs were presented for 6 s and separated by a jittered fixation period (2 s to 6 s) with a mean of 4 s.

During associative recognition task runs (Figure 1), participants viewed face-belief pairs from the previous ABT run. Of these, 20 pairs (two-thirds) were in identical format and 10 pairs (one-third) were in rearranged format. Participants were instructed to judge whether each pair was intact or recombined (memory probe), in addition to providing a corresponding confidence rating (confidence probe). Participants were instructed to respond 'intact' if they judged a given face-belief pair to be unchanged from the immediately preceding run and were instructed to respond 'recombined' if they believed the face and statement had not appeared together at encoding. Each memory probe trial was presented for 4 s and was immediately followed (with no jittered fixation period) by a confidence probe trial of 4 s during which participants made a subjective confidence rating (1 = very low, 4 = very high) for their intact or recombined decision. Each compound retrieval trial (memory probe and confidence probe) was separated by a jittered fixation period (2 s to 6 s) with a mean of 4 s (see Figure 1). The faces and statements presented in each associative recognition run were always drawn from the immediately preceding ABT run. The order in which each pair of runs (ABT and corresponding associative recognition) appeared was counterbalanced across participants, while the order of trial presentation within runs was random.

Because a main goal of the present investigation was to determine the influence of personal beliefs on subsequent associative recognition memory, trials during ABT runs were classified according to both ABT response and subsequent memory response. Agreement rating responses were analyzed in two ways: (a) agreement intensity (i.e. arousal) was measured by comparing high-intensity agreement (SA/SD collapsed) vs. low-intensity agreement (WA/WD collapsed), and (b) valence was measured by comparing agree (SA/WA collapsed) vs. disagree (SD/WD collapsed). With respect to subsequent memory, responses

to intact trials on the associative recognition memory task runs were used to back-sort ABT trials, allowing us to examine neural activity associated with trials that were either subsequently remembered (correctly called intact) or subsequently forgotten (incorrectly called recombined) at different levels of confidence. The present investigation focuses exclusively on neural activity collected during the ABT (i.e. encoding) runs and does not address imaging data collected during the associative recognition memory task (i.e. retrieval) runs.

To calculate a corrected recognition score (hit-false alarm), recombined trials incorrectly classified as intact were classified as false alarms. To ensure no instances of either 100% or 0% hit and false alarm rates, we used a procedure recommended by several studies (Quamme, Yonelinas, & Norman, 2007; Snodgrass & Corwin, 1988): hit rate was computed as: [(#Hits+0.5)/(total number of intact items+1)] and false alarm rates computed as: [(#FA +0.5)/(total number of recombined items+1)].

fMRI Methods

All MRI data acquisition was conducted with a 3-T GE scanner. Scanner noise was reduced with earplugs, and head motion was minimized with foam pads. Stimuli presented across the twelve functional runs were projected onto a mirror located at the rear of the scanner. Behavioral responses were recorded with two 4-key fiber-optic response boxes placed on the left and right hands respectively (Resonance Technology, Inc.), and when necessary, vision was corrected using MRI-compatible lenses that matched the distance prescription used by the participant. High-resolution T1-weighted structural images were collected using a 3D, T1-weighted FSPGR sequence (256×256 matrix, 96 slices, and 1.9 mm slice thickness). Functional images were acquired using a SENSE inverse-spiral sequence (64×64 matrix, TR=2000 ms, TE= 30ms, FOV = 24 cm, flip angle = 70°). Thirty-four contiguous slices were acquired in an interleaved fashion. Slice thickness was 3.8 mm, resulting in $3.75 \times 3.75 \times 3.8$mm voxels.

Preprocessing and data analyses were performed using SPM5 software implemented in Matlab (www.fil.ion.ucl.ac.ul/spm/). After discarding the first five volumes of each run, functional images were corrected for slice time acquisition and motion. These images were then spatially normalized into the Montreal Neurological Institute (MNI) template and spatially smoothed using a Gaussian kernel of 8 mm FWHM. Two first-level models were constructed, one examining ABT responses collapsing across memory (i.e. irrespective of subsequent score on the associative recognition phase), and the second examining encoding processes with subsequent memory included as a parametric modulator. In each model, the evoked hemodynamic responses was modeled with a delta (stick) function corresponding to the onset of stimulus presentation convolved with a canonical hemodynamic response function in the context of the general linear model (GLM). Regressors for head motion and run mean were also included in each of the models, which differed only slightly in their designs.

The first model included 4 conditions of interest corresponding to each level of the ABT scale: 1 = strongly disagree (SD), 2 = weakly disagree (WD), 3 = weakly agree (WA), and 4 = strongly agree (SA), with an additional regressor for trials with no response or a 'don't

know' response. In the non-memory model, ABT regressors also included trials that were subsequently recombined. The second memory-related model had the same structure but included one additional regressor for these subsequently recombined trials, which had no memory score at encoding. For each of the 4 ABT conditions of interest, a linear parametric memory modulator was included, where values corresponded to subsequent memory outcome. To account for variability in the use of the 4-pt memory confidence scale across participants, three modulator levels were used, which corresponded to: 1) subsequent misses of any confidence rating 2) subsequent low-confidence hits (confidence rating values 1 and 2) and 3) subsequent high confidence hits (confidence rating values 3 and 4). For each of the models, individual subject contrasts for each of the 4 conditions of interest were generated at the fixed effects level. For each model, within-subjects contrasts were then generated for the main effects of arousal (strong vs. weak) and valence (agree vs. disagree), as well as for the arousal by valence interaction. For the first model, these contrasts corresponded to the overall effect of the ABT judgments collapsing across memory. In the second model, contrasts reflected a comparison of the parametric subsequent memory effect in each condition.

Planned contrasts were used to evaluate these main effects and interactions by submitting each within-subjects contrasts to a one sample t-test. Monte Carlo simulations were obtained using 3dClustSim (compiled January 2017) from the AFNI software package (Ward, 2000). With an uncorrected height threshold of $p < 0.001$, the $p < 0.05$ corrected cluster extent for the whole brain mask was 26 voxels. Given our hypotheses about the role of the amygdala in arousal, the threshold for the non-memory contrast of arousal (strong vs. weak) was $p < 0.005$ with a corrected cluster extent of three voxels calculated for an anatomical ROI of bilateral amygdala from the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). Given the small size of the amygdala mask, we also ran a confirmatory test of the arousal contrast across all voxels in the anatomical mask with an uncorrected threshold of $p < 0.05$.

Results

Behavioral results

Behavioral normative study—Our *first goal* was to test a paradigm for investigating belief-related emotions that was ecologically-valid, feasible for fMRI, and also internally and externally valid. In addition, the paradigm was designed to elicit substantial emotional arousal and produce differences in subsequent memory. To assess internal validity we measured the correlation between subjective agreement ratings made for originally endorsed viewpoints (e.g., '*Thinks the law should limit experiments with human embryos*') and companion counter-viewpoints for each belief (e.g. '*Supports increased funding for stem cell research*'). Ratings for opposing versions of each belief were first correlated within subject and then averaged together to form a group correlation. A one-sample t-test was then performed to determine whether the group correlation was significantly different from a chance level of zero. Supporting the internal validity of the belief stimuli, we found a negative correlation (mean within-subject $r = -0.35$, SD = 0.11) between the original and counter-viewpoint versions of each belief, which was significantly different from chance

(95% CI: −0.40 to −0.30); t(20) = −14.58, $p_{(two-tailed)}$ < 0.001). A significant negative correlation (r = −0.45, SD = 0.19; 95% CI: −0.53 to −0.36; t(21) = −10.83, $p_{(two-tailed)}$ < 0.001) was also found when evaluating the subjective agreement ratings provided by those participating in the imaging study.

To assess external validity, we compared the subjective agreement ratings for participants in the normative study that were self-identified liberals (n = 14) or conservatives (n = 7). Although these groups were not matched in size, Levene's Test for Equality of Variances was not significant and an independent sample t-test was performed. For belief statements traditionally endorsing a liberal stance, self-identified liberals had significantly higher subjective agreement ratings (M = 3.04, SD = 0.20) compared with self-identified conservatives (M = 2.61, SD = 0.28), t(19) = 3.967, $p_{(two-tailed)}$ = 0.001. By contrast, for belief statements traditionally endorsing a conservative stance, self-identified conservatives had significantly higher subjective agreement ratings (M = 2.14, SD = 0.21) compared with self-identified liberals (M = 1.82, SD = 0.281), t(19) = −2.64, $p_{(two-tailed)}$ = 0.016. We also compared subjective agreement ratings for participants in the imaging study that were self-identified liberals (n = 11) or conservatives (n = 4). The number of participants reflects those who completed the online questionnaire assessing political affiliation. An independent sample t-test revealed that for belief statements endorsing a liberal stance, self-identified liberals had significantly higher subjective agreement ratings (M = 3.23, SD = 0.29) compared with self-identified conservatives (M = 2.62, SD = 0.25), t(13) = 3.67, $p_{(two-tailed)}$ = 0.003. By contrast, for belief statements endorsing a conservative stance, self- identified conservatives had significantly higher subjective agreement ratings (M = 2.36, SD = 0.20) compared with self-identified liberals (M = 1.73, SD = 0.19; t(13) = −5.52, $p_{(two-tailed)}$ < 0.001). While we acknowledge that ratings provided by self-identified conservatives for both liberal and conservative stances do not differ greatly, it is important to consider that there were very few self-identified conservatives in our sample (N = 4), with all identifying themselves as merely 'somewhat conservative.' Testing a larger number of self-identified conservatives who strongly identify with this political ideology may be fruitful in terms of further testing the external validity of the ABT.

Lastly, we investigated whether or not the ABT can elicit significant emotional arousal. Participants in the normative study rated their emotional arousal (from 1 = "does not move me" to 4 = "moves me quite a bit"). We measured emotional arousal by comparing high arousal (SA/SD) to low arousal (WA/WD) judgments. Mean emotional arousal ratings for strongly disagree (SD), weakly disagree (WD), weakly agree (WA), and strongly agree (SA) judgments are plotted in Figure 2-A. A 2 (Arousal: Strong vs. Weak) × 2 (Valence: Disagree vs. Agree) repeated measures ANOVA showed that the most pronounced differences in subjective ratings were related to agreement intensity (Strong vs. Weak), which showed a significant main effect [F(1, 20) = 84.79, p < 0.001], with pairwise LSD contrasts confirming significantly higher emotional arousal ratings for SD than for WD (p < 0.001) and for SA than WA (p < 0.001). The effect of valence was more subtle than that of arousal, but also showed a significant main effect [F(1, 20) = 6.62, p = 0.018], and the interaction was also significant [F(1, 20) = 6.16, p = 0.022]. Thus, we found that ABT responses elicited substantial emotional arousal, particularly in instances where participants reported strong vs. weak agreement or disagreement with the statements.

In-scanner behavior—Responses for the scanned ABT task were well distributed, with a similar proportion (with SEM) of trials falling into each of the 4 main conditions [SD: 24.1% (5.1%); WD: 24.3% (5.1%); WA: 23.9% (5.1%); SA: 23.2% (4.9%); Don't Know: 4.5% (1.0%)]. We focused on response times (RTs), which represent the main behavioral measure given that the ABT responses are neither correct nor incorrect. As illustrated by Figure 2-B, when the 4 conditions are ordered from SD to SA, RTs show a clear inverted U-shape function. A 2 (Arousal: Strong vs. Weak) $\times$ 2 (Valence: Disagree vs. Agree) repeated measures ANOVA on mean RTs for each condition confirmed a main effect of Arousal [$F(1, 21) = 90.75$, $p < 0.001$], with pairwise LSD tests showing significantly faster RTs for SD vs. WD ($p < 0.001$) and for SA vs. WA ($p < 0.001$) trials. The main effect of Valence was less pronounced, but was also significant [$F(1, 21) = 8.35$, $p = 0.009$] although there was no significant interaction [$F(1, 21) = 0.04$, $p = 0.839$]. The finding of faster responses for SD/SA than WD/WA trials broadly matches the pattern of responses seen for subjective emotional arousal in the normative study; both of these measures were most strongly influenced by the intensity (SD/SA vs WD/WA) of participant ratings.

During the retrieval phase, the proportion of intact face-statement pairs correctly identified as intact was first calculated by collapsing across confidence, which yielded an overall hit rate of M = 0.73; SEM = 0.023, and a false alarm rate (proportion of recombined trials incorrectly deemed intact) of M = 0.21; SEM = 0.035. The mean proportion (and SEM) of responses for a given confidence level was then examined within the set of hits [level 1: 0.12 (0.029); level 2: 0.16, (0.018); level 3: 0.22 (0.020); level 4: 0.43 (0.041)]. Because they were more numerous than false alarms, a similar memory confidence breakdown was also calculated for correct rejections [level 1: 0.10, (0.024); level 2: 0.18 (0.017); level 3: 0.21 (0.022); level 4: 0.43 (0.041)]. Of central interest for neuroimaging analysis was how the arousal dimension of ABT decisions might influence later memory for intact trials. Confirming the memory-related dimension of ABT judgments, a 2 (Arousal: Strong vs. Weak) $\times$ 2 (Valence: Disagree vs. Agree) repeated measures ANOVA on overall hit rates revealed a significant main effect of Arousal [$F(1, 21) = 11.38$, $p = 0.003$] with pairwise LSD tests confirming significantly higher overall hit-rates (collapsed across all confidence levels) for SD than WD trials ($p = 0.011$) and for SA compared to WA trials ($p = 0.020$). Neither the main effect of Valence [$F(1, 21) = 0.002$, $p = 0.970$] nor the interaction [$F(1, 21) = 0.073$, $p = 0.790$)] were significant.

The effects of ABT judgments on associative recognition were particularly clear when focusing on the most confident recognition responses ('very high' confidence; level 4). As illustrated in Figure 2-C, the proportion of very high- confidence hits shows a clear U-shaped function reflecting better memory for strong judgments (SD/SA) than for weak judgments (WD/WA). Results of a two-way repeated measures ANOVA with factors of Arousal and Valence revealed a significant main effect of emotional arousal on high-confidence hit rates [$F(1, 21) = 17.02$, $p < 0.001$]; however, there was no significant main effect of valence [$F(1, 21) = 0.30$, $p = 0.59$]. The interaction showed a trend towards significance [$F(1, 21) = 4.20$, $p = 0.053$], with a greater difference in very high confidence hits between strong vs. weak disagree comparisons relative to strong vs. weak agree comparisons. One potential caveat in the memory results is the interleaved encoding-

retrieval block structure, which was designed to achieve a reasonable level of hits and misses. Overall memory discriminability did not change across runs, as evidence by a 1×6 repeated measures ANOVA on d-prime [$F(5, 105) = 0.99$, $p = 0.428$; linear test: F $(1, 21) = 0.36$, $p = 0.556$]. However, while a corresponding test of bias showed no large differences between runs [$F(5, 105) = 1.96$, p = 0.09], a small but significant increase in bias across the study was evident from a test of the linear trend [linear test: $F(1,21) = 5.55$, p = 0.028]. Despite this, the overall influence of belief-related emotions in the ABT was found to have a robust impact on subsequent memory and this impact was mediated by arousal (strong vs. weak) rather than valence (agree vs. disagree). Furthermore, better memory for strong (SD/SA) judgments cannot be attributed to longer encoding time because these judgments had faster RTs than weak (WD/WA) judgments (see Figure 2-B). For the same reason, larger SMEs for strong vs. weak judgments cannot be attributed to differences in time-on-task.

fMRI results

As described above, we analyzed fMRI data as a function of participant ratings during the ABT task. Initially, we looked at ABT responses collapsed across memory and examined contrasts showing a main effect of arousal (SD/SA vs. WD/WA judgments) or valence (SA/WA vs. SD/WD judgments) as well as the interaction. We then examined subsequent memory effects for these conditions by modulating each trail type by a parametric regressor coding for subsequent memory score. Below, we consider brain activity associated with arousal, valence, and the impact of these factors on subsequent memory.

Our *second goal* was to investigate the neural mechanisms of belief-related emotional arousal, which we identified by comparing activity for SD/WD judgments to activity for WD/SA judgments. Given that we had a strong prediction about the amygdala, we performed an ROI analysis on this region. As illustrated by Figure 3, the results confirmed our prediction: clusters in both the left and right amygdala showed greater activity for strong (SD/SA) than weak (WD/WA) judgments, with more extensive effects in the right amygdala (see Figure 3). In addition to this cluster-based analysis, a confirmatory test showed a significant difference between strong and weak agreement conditions based on the average activity across all voxels in the bilateral anatomical amygdala mask [$t(21) = 2.28$, p = 0.017]. After confirming this prediction, we performed a whole-brain analysis to identify other brain regions showing emotional arousal effects, as well as those that showed effects of valence or an arousal by valence interaction. As listed in Table 1, regions showing greater activity for strong (SD/SA) than for weak (WD/WA) judgments included medial PFC (largely within the anterior cingulate cortex/ACC), ventrolateral PFC (right inferior frontal gyrus), and right superior temporal gyrus. For the reverse pattern of greater activity for weak (WD/WA) than strong (SD/SA) judgments, no clusters survived the corrected threshold.

In examining the main effect of agreement, the contrast of agree vs. disagree revealed a region in occipital cortex with higher activity for agree than disagree trials, with additional clusters showing the same direction of effects in ACC and orbitofrontal cortex (Table 1). No regions showed significantly greater activity for disagree than agree, while tests for interactions showed only a single region in precentral gyrus (SA > WA vs. SD > WD).

Finally, we turned to our third and most important goal, which was to investigate the neural mechanisms of the enhancing effect of belief-related emotion on subsequent memory. We compared differences in the size of SMEs (defined by the three-point linear parametric of subsequent memory score) as a function of arousal and valence. To this end, we found an emotional arousal (SD/SA > WD/WA) × subsequent memory (positive increase) interaction in a single region of mPFC (maxima: $x = 4$, $y = 49$, $z = -11$) (see Figure 4-A; Table 1). As illustrated by Figure 4-B, this region showed larger SMEs for strong (SD/SA) than weak (WD/WA) judgments both for disagree and agree, with a numerically larger difference in the former. No brain region showed the reverse effect of larger SMEs for weak (WD/WA) than strong (SD/SD) judgments. Additionally, we found no brain regions showing memory-modulated main effects of valence, or memory-modulated arousal by valence interactions. This finding suggests that the impact of belief-related emotion on subsequent memory reflects emotional arousal rather than valence. Interestingly, the amygdala did not show arousal-related subsequent memory effects, and no other regions showed an overall main effect of subsequent memory at the cluster corrected threshold. Although regions like the MTL often appear in general contrasts of successful episodic encoding (Kim, 2011), the frontal location of subsequent memory effects here matches findings from related work in which the social aspect of stimulus or task-orientation was explicitly manipulated (Harvey et al., 2007; Mitchell et al., 2004). Furthermore, while the amygdala has often been implicated in memory enhancing effects of emotion for basic emotional stimuli (e.g., affective pictures), the present results suggest that the mPFC may be pivotal for subsequent memory when emotion relates to personal beliefs.

Discussion

Whereas the vast majority of behavioral and fMRI studies on emotional memory have focused on emotions not mediated by complex personal beliefs, we investigated emotions closely linked to the individual opinions of each participant. The current study had three main goals. Our first goal was to test a paradigm that, in addition to being ecologically-valid and scanner-friendly, would be internally and externally valid, would elicit significant emotional arousal, and would enhance subsequent memory. Assuming this paradigm would generate significant emotional arousal, our second goal was to identify the neural mechanisms mediating arousal effects. Finally and most importantly, our third goal was to explore the neural correlates of emotion-related subsequent memory effects.

The study yielded three main results. Fulfilling our first goal, the *agreement with beliefs task* (ABT) proved both internally and externally valid, elicited significant emotional arousal (Figure 2-A), and showed memory-related differences based on participant responses (Figure 2-C). In pursuit of our second goal, we observed a positive relationship between increased emotional arousal during the ABT and activation of the amygdala (Figure 3). This finding indicates that the amygdala, a region strongly associated with emotional arousal driven by more basic stimuli, also mediates emotional arousal related to personal beliefs. Finally, achieving our third goal, we linked the subsequent memory-enhancing effects of emotional arousal in the ABT to the mPFC (Figure 4). This finding indicates that emotions linked to evaluation of personal beliefs enhance memory via the mPFC, a region strongly

associated with self-referential processing. The three main results are discussed in separate sections below.

The ABT: Contributing to fMRI paradigms investigating emotional memory

As noted in the introduction, the majority of prior studies on the neural mechanisms of emotional evaluation and memory have ignored emotion related to personal beliefs. Examples of the few studies that have employed such emotions include studies in which participants provided ratings of how much emotion they felt when viewing faces of politicians who share their ideology versus those that do not (Kaplan, Freedman, & Iacoboni, 2007), or rated the reasonability of statements related to their particular ethnic group (Bruneau & Saxe, 2010). In general, these studies have focused mostly on social questions, such as the formation of impressions or the role of ethnic or political identity, rather than on the generation of emotional arousal (but see Gozzi et al., 2010). As noted before, emotional arousal elicited from social situations is difficult to study in the laboratory because the same event can elicit heterogeneous responses across individuals. To address this issue, the ABT task measured emotional arousal by asking participants to rate the extent to which they agreed or disagreed with socio-political issue statements, under the assumption that emotional arousal would relate to the strength of agreement or disagreement. Consistent with this assumption, a normative study showed that higher agreement intensity (strong agree/disagree judgments) was associated with higher subjective emotional arousal ratings in comparison to weak agree/disagree judgments (Figure 2-A).

To validate our assumptions that the ABT could be used to investigate emotions generated from abstract stimuli, we carried out tests to measure both internal and external validity. As a test of internal validity, we examined whether the belief statements used in the study were truly tapping into opposing perspectives of a socio-political issue. Among participants in the normative and imaging studies, correlation analyses revealed negative relationships between agreement ratings of opposing viewpoints of the same issue. Thus, the statements adequately conveyed different perspectives of a particular socio-political issue, and participants were largely consistent with their agreement ratings through the course of the study. As a test of external validity, we examined the relationship between the self-reported political ideology of participants and the subjective agreement ratings made in response to liberal or conservative belief statements. In comparing the mean ratings provided by both groups, we found that for liberal beliefs, self-identified liberals provided higher agreement ratings than self-identified conservatives, with the reverse effect for ratings provided for conservative beliefs. While many of these findings are intuitive, they nonetheless support the notion that the beliefs used in the ABT accurately capture viewpoints across a wide range of issues that are relevant to individuals who identify with either liberal or conservative political ideologies. Taken together, these findings indicate that the ABT can contribute to the existing set of fMRI paradigms investigating emotional memory, which are dominated by the use of stimuli that have more homogenous responses across individuals.

Neural correlates of emotional arousal generated by evaluation of personal beliefs

Our *second goal* was to investigate the neural mechanisms of emotional arousal triggered by social situations, which we identified by comparing activity for high arousal (SA/SD)

judgments vs. low arousal (WA/WD) judgments. Given a hypothesized effect in the amygdala, we focused on this region as an ROI. Confirming our hypothesis, amygdala activity was greater for SA/SD than for WA/WD judgments (Figure 3). This finding generalizes the link between the amygdala and emotional arousal found elsewhere in the literature (Cahill & McGaugh, 1998; Hamann, 2001; LaBar & Cabeza, 2006) to the domain of emotions triggered by personal beliefs. Whereas many studies of emotion expect to elicit similar responses to the same stimuli across participants (e.g., pictures of dangerous animals), in the case of the ABT, arousal-related amygdala responses are generated by the comparison of pre-existent personal beliefs to an external statement. The beliefs that elicited SD judgments in some participants, elicited WD or WA judgments in other participants, and vice-versa. Thus, in the ABT, amygdala activity did not track properties of the stimuli themselves but instead related to how strongly belief statements matched each participant's own viewpoint.

Although increases in amygdala activity are often associated with aversive or fearful stimuli (Phan, Wager, Taylor, & Liberzon, 2002), a number of studies have also found that positively valenced stimuli elicit amygdala responses. Along the valence dimension, the non-linear profile of amygdala activity in the present task mirrors that of several studies exploring different dimensions of face processing. For example, while faces perceived as particularly untrustworthy were associated with increased activity in bilateral amygdala, a similar increase was also found in response to faces with the highest (vs. intermediate) trustworthiness ratings (Said, Baron, & Todorov, 2009). A similar non-linear response in the amygdala was also found for faces with high or low attractiveness (Winston, O'Doherty, Kilner, Perrett, & Dolan, 2007), and other work has shown that the amygdala is sensitive to general face salience (Santos, Mier, Kirsch, & Meyer-Lindenberg, 2011) or distinctiveness (Said, Dotsch, & Todorov, 2010). Along these lines, the observed pattern of amygdala activity during the ABT may partly reflect the subjective salience of the various statement topics, which is reflected in the strength of agreement ratings. One obvious difference with these other studies is that, although aspects like attractiveness and trustworthiness have a large subjective component, views about political statements ("Thinks democracy cannot work in every country" or "Supports campaign finance reform") require more involved extraction of meaning, followed by evaluation of conceptually complex personal opinions.

To explore other regions that might be involved in the arousal dimension of ABT judgments, we performed a whole brain contrast and found greater activity for strong (SA/SD) vs. weak (WA/WD) judgments in medial/ventrolateral regions of PFC as well as in posterior superior temporal sulcus and areas of rhinal cortex adjacent to the amygdala (Table 1). These frontal and MTL regions have been associated with episodic memory retrieval (Spaniol et al., 2009), and one possibility is that high-arousal judgments involve the retrieval of personal memories related to the issue at question. For example, while assessing one's personal view on the use of human embryos, a participant could try to remember how they felt the last time they heard an alternative view being espoused on television, or in a personal interaction. Remembering a strong negative feeling might therefore coincide with a "strongly disagree" judgment, whereas issues eliciting intermediate judgments might be connected to fewer past experiences, limiting the influence of mnemonic retrieval in ABT judgments. Thus, although emotional arousal elicited by evaluating personal beliefs appears to share amygdala-related

components with emotional arousal elicited by other types of stimuli, it likely involves additional processes that may include episodic retrieval.

In addition to potential retrieval-related processes, strong agreement/disagreement judgments in the ABT may also reflect an overall increase in self-referential processing, which is often associated with mPFC engagement (Denny, Kober, Wager, & Ochsner, 2012). This possibility relates closely to the notion that a participant's decision on a given trial is influenced by the subjective salience of the statement. Both agree and disagree trials characterized by strong responses are likely to involve increased access to personal beliefs and increased focus on self-identity. Just as past studies have shown that accessing knowledge about the self involves mPFC (e.g., D'Argembeau et al., 2007; Kelley et al., 2002b), the emergence of this region in the present arousal-related contrast may reflect the necessity of comparing the semantic content of statements with personal beliefs before responses can be made. In general, the finding of additional arousal components for emotions generated from abstract, complex stimuli makes intuitive sense given that such emotions are less automatic and involve greater elaboration (Knutson, Wood, Spampinato, & Grafman, 2006; Sakaki, Niki, & Mather, 2012). Ultimately, deciding how one feels about complex issues such as embryo usage or privacy is likely to involve a more complex network in which additional interpretation of stimuli is required before affective significance takes shape.

Although our focus was on agreement intensity, we performed a whole-brain contrast for regions showing differences in agreement valence. While no regions showed stronger activity for disagree than agree trials, an area of lateral orbitofrontal cortex along with a region in ACC showed the reverse pattern. While several studies across a range of domains have reported activations for positive vs. negative stimuli in orbital (Small et al., 2003; Tsukiura & Cabeza, 2011) or medial (Dolcos et al., 2004a) PFC, it is unclear exactly how these stimulus-driven valence effects might relate to statements expressing a viewpoint with which one agrees rather than disagrees. Meta-analyses of other literature (Denny et al., 2012; Northoff et al., 2006) have shown that various regions in frontal midline are consistently involved in tasks requiring comparisons between oneself and others. For example, the peak coordinate in ACC for the current agree vs. disagree contrast falls near that of a similar region (MNI: 18 57 9) found to be sensitive to comparisons between similar vs. dissimilar others (Mitchell et al., 2006). This convergence makes sense, given that the self-similarity measures used by Mitchell and colleagues were based in part on shared political opinions. Although perceived similarity with the individuals accompanying ABT statements was not probed directly, it seems reasonable to assume that agreement decisions reflect a rough index of perceived self-similarity, at least along the particular dimension expressed in the statement. However, the medial frontal region sensitive to valence also extends more dorsally, and other work suggests that regions of dorsal ACC are sensitive to tasks requiring a range of social comparisons involving the self (Denny et al., 2012; Northoff et al., 2006), including graded measurements reflecting how subjects identify with the actions attributed to other individuals (Leshikar et al., 2016). While the ABT differs in many respects from tasks requiring an explicit comparison of self-similarity (based on shared traits or activities), the limited valence results suggest common elements with such tasks. Future research will

be useful in characterizing how agreement decisions relate to non-social emotional valence effects and to the various factors that influence judgments of social similarity.

In sum, our initial set of questions yielded several results related to emotional arousal. Findings from the normative study showed that ABT decisions were in fact associated with subjective emotional arousal. Neuroimaging contrasts showed that the amygdala was sensitive to the arousal dimension of ABT responses. However, additional frontal and MTL regions also showed increased activity for high arousal judgments, consistent with the operation of additional mnemonic and self-referential processes that may have coincided with arousal differences in the ABT.

Neural mechanisms of arousal-related enhancements on subsequent memory

Our third and most important goal was to investigate the neural mechanisms supporting memory enhancement linked to belief-related arousal. Compared to arousal (WA/WD) judgments, high arousal (SA/SD) judgments yielded a larger subsequent memory effect (SME) in the mPFC (Figure 4-B). This was the only region to show an emotional arousal effect on subsequent memory, and no areas were found for the opposite effect (greater SMEs WD/WA vs. SD/SA judgments).

The behavioral finding that memory is increased for ABT trials with higher arousal shows that the memory-enhancing power of emotions is not limited to lower-level affective features and can be triggered by evaluation of complex personal beliefs. It should not surprising that memory systems evolved to preferentially encode events associated with negative or positive emotions in the context of basic aversive or appetitive stimuli, where relevance to survival is clear (e.g., the fear of encountering a predator; the happiness of finding food). Indeed, an evolutionary account for such mnemonic benefits has been discussed in the emotional memory literature (Adolphs, Russell, & Tranel, 1999; Hamann, 2001). An account of preferential memory for belief-related emotions is less obvious. One possibility is that these emotions are indirectly related to survival through interactions with surrounding social groups. In the case of the ABT, remembering who agrees with us and who disagrees with us could be essential for distinguishing between friends (close to the self) and foes (others) and/or defining our position within the group. Given the scarcity of research on emotions generated from evaluating personal beliefs, it is difficult at present to address the basic question of *why* these emotions enhance memory.

Turning to the more tractable question of *how* such emotions enhance memory, the current study showed that faces paired with arousing beliefs elicited larger SMEs in mPFC (Figure 4-B). It is interesting that the memory-enhancing effect of emotional arousal was not found in the amygdala, which has often been reported in studies of emotional memory (Cahill & McGaugh, 1998; Hamann, 2001; LaBar & Cabeza, 2006). This current finding is not wholly unexpected, given that past work from our lab and others has shown that differential processing across a consistent set of emotional stimuli can modulate the emotion-related encoding mechanisms (Pais-Vieira, Wing, & Cabeza, 2016; Ritchey, LaBar, & Cabeza, 2011; Talmi, Anderson, Riggs, Caplan, & Moscovitch, 2008). Without a direct contrast between emotions generated automatically or those linked to further elaboration, it is unclear whether the lack of emotional arousal-related memory effects in the amygdala is a

fundamental difference between these two types of emotions. Given the methodological differences between the current study and typical fMRI studies using basic emotional stimuli, future work would benefit from direct comparisons between different types of emotion inducing stimuli within the same subjects.

Importantly, the lack of memory-enhancing effects in the amygdala cannot be attributed to a weak emotional arousal manipulation or to low fMRI signal in this region given that the amygdala showed a significant activity difference between strong (SD/SA) and weak (WD/WA) trials independently of memory (Figure 3-B). One possible explanation of why the amygdala did not mediate the impact of arousal on memory concerns the associative recognition format of the memory test. Most studies linking the amygdala to enhanced memory formation have focused on memory of items (e.g., words, pictures) rather than on associations between different items, such as faces and beliefs. According to one view, arousal (as elicited by survival-related emotional stimuli) can enhance binding between an object and its features (e.g., color) but not between an object and other distinct objects or background contextual information (Mather, 2007). Consistent with this theory, there is evidence that arousal-related amygdala activity enhances the vividness of subsequent item recognition, but not context memory (Dougal, Phelps, & Davachi, 2007; Kensinger & Schacter, 2006). In contrast, arousal-related mPFC activity in the current study was associated with better associative recognition for face-belief pairs.

The finding that arousal-related mPFC activity enhanced subsequent memory aligns with and extends many results reported across several lines of ongoing research. Specifically, our finding is consistent with results from the self-referential processing literature, which often uses trait adjective paradigms in which people describe the self-relevance of different traits or descriptions (Heatherton et al., 2006; Kelley et al., 2002). For example, Macrae et al. (2004) found an mPFC region that showed greater activity for self-relevant vs. non-relevant adjectives and for subsequently remembered vs. forgotten adjectives (see also Benoit, Gilbert, Volle, & Burgess, 2010; Gutchess, Kensinger, & Schacter, 2010). Our findings extend this result by showing that the relation between self-relevance and subsequent memory in mPFC is particularly sensitive to strength of agreement, and shows a non-linear pattern with respect to agreement valence. (Figure 4-B). In related work explicitly examining source memory, Leshikar and Duarte (2012) found that mPFC activity predicted subsequent memory for object-scene pairs when participants focused on a self-related dimension (pleasantness) rather than color (Leshikar & Duarte, 2012). Our finding, based on associative recognition performance, extends this result by showing that even within tasks like the ABT that always contain a self-related dimension, subjective appraisals (e.g. agreement intensity) can further influence mPFC contributions to relational memory.

The observed involvement of mPFC in memory formation during the ABT is also consistent with work on impression formation, which explores how processing traits or actions related to individuals shapes the development of social opinions (e.g. Cassidy, Leshikar, Shih, Aizenman, & Gutchess, 2013; Gilron & Gutchess, 2012; Mitchell, Banaji, & Macrae, 2005). In one study, Mitchell et al. (2004) had participants associate faces with actions reflecting a personality trait (e.g., *He refused to loan his extra blanket to the other campers* for the "inconsiderate" trait) while focusing either on forming an impression about the person or on

the temporal sequence of trial presentation (Mitchell et al., 2004). When focusing on forming an impression, activity in dorsal mPFC region predicted associative memory for face-sentence pairs, while the hippocampus supported subsequent memory during the temporal sequence task. In fact, while hippocampal involvement in subsequent memory is evident in specific conditions within some impression formation studies (Leshikar et al., 2016), medial frontal SMEs often appear in the absence of corresponding hippocampal SMEs (Gilron & Gutchess, 2012; Macrae et al., 2004). Similarly, mPFC encoding effects have also been found during the processing of social vs. non-social pictures, with only the latter condition showing hippocampal SMEs (Harvey et al., 2007). Thus, with mPFC as the only region supporting subsequent memory for strong vs. weak ABT judgments, the current results strengthen the existing link between encoding success and medial frontal areas during tasks that involve self-reference and social processing.

Beyond the role of the mPFC in self-referential processing, associative memory, and impression formation, the current finding is also related to a distinct literature regarding the contributions of mPFC to processing pre-existing knowledge or "schemas" (Knutson et al., 2006; van Kesteren et al., 2013; van Kesteren, Rijpkema, Ruiter, & Fernández, 2010; van Kesteren et al., 2012). Citing rodent evidence (Tse et al., 2007, 2011), a recent theory postulates that information congruent with pre-existing schemas is encoded primarily by mPFC, with incongruent information encoded primarily by the MTL (van Kesteren et al., 2012). Consistent with this theory, greater SMEs in mPFC were found for encoding congruent object-scene pairs (e.g., tennis court – tennis racket) than for incongruent pairs (e.g., tennis court – umbrella), whereas MTL showed the opposite effect (van Kesteren et al., 2013). Initially, the current findings seem inconsistent with this theory because, compared to intermediate beliefs, we found larger SMEs in mPFC for beliefs that were highly congruent with the self and for beliefs that were highly incongruent with the self. However, this finding can be harmonized with the schema theory given that schemas about political issues include the position one supports and the position one opposes. In this sense, having a schema about an issue would be equivalent to having a strong opinion about the issue, which is required for both SA and SD judgments. This interpretation fits with evidence from our normative study, where issue knowledge was higher for statements generating strong vs. weak agreement.

Conclusions

In order to address a relatively neglected area of the emotional memory literature, we tested a paradigm investigating emotion triggered from personal beliefs: the Agreement with Beliefs Task (ABT). The present paradigm provided measures of agreement intensity or arousal (strong vs. weak judgments) and valence (agree vs. disagree judgments) and revealed how these two factors affect subsequent associative recognition memory. The study yielded three main findings. First, behavioral results showed that the ABT is internally and externally valid and that SA/SD judgments elicited greater subjective measures of emotional arousal and better subsequent memory than WA/WD judgments. Second, fMRI analyses of the judgment phase showed that emotional arousal in the ABT (SA/SD vs. WA/WD) was associated with greater activity in the amygdala, as well regions of PFC. This finding indicates shared emotional arousal mechanisms for social- and non-social triggers of affect

and suggests additional processes involved in arousal-related judgments for complex stimuli. Finally, and most importantly to the present investigation, subsequent memory analyses showed that arousal yielded SMEs in the mPFC but not in the amygdala. This finding is consistent with evidence linking the mPFC to self-referential processing, subsequent associative memory formation, and an increasingly apparent role in the encoding of socially relevant information.

Acknowledgments

References

Adolphs R, Russell JA, Tranel D. A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. Psychological Science. 1999; 10(2):167–171. http://doi.org/10.1111/1467-9280.00126.

Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience. 2006; 7(4):268–77. http://doi.org/10.1038/nrn1884. [PubMed: 16552413]

Benoit RG, Gilbert SJ, Volle E, Burgess PW. When I think about me and simulate you: medial rostral prefrontal cortex and self-referential processes. NeuroImage. 2010; 50(3):1340–9. http://doi.org/10.1016/j.neuroimage.2009.12.091. [PubMed: 20045478]

Bentley SV, Greenaway KH, Haslam SA. An online paradigm for exploring the self-reference effect. PLOS ONE. 2017; 12(5):e0176611. http://doi.org/10.1371/journal.pone.0176611. [PubMed: 28472160]

Bradley SD, Angelini JR, Lee S. Psychophysiological and memory effects of negative political ADS: Aversive, arousing, and well remembered. Journal of Advertising. 2007; 36(4):115–127. http://doi.org/10.2753/JOA0091-3367360409.

Brod G, Lindenberger U, Wagner AD, Shing YL. Knowledge Acquisition during Exam Preparation Improves Memory and Modulates Memory Formation. The Journal of Neuroscience. 2016; 36(31):8103–11. http://doi.org/10.1523/JNEUROSCI.0045-16.2016. [PubMed: 27488631]

Bruneau EG, Saxe R. Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. NeuroImage. 2010; 52(4):1704–11. http://doi.org/10.1016/j.neuroimage.2010.05.057. [PubMed: 20510370]

Cahill L, McGaugh J. A novel demonstration of enhanced memory associated with emotional arousal. Consciousness and Cognition. 1995; 4:410–421. [PubMed: 8750416]

Cahill L, McGaugh JL. Mechanisms of emotional arousal and lasting declarative memory. Trends in Neurosciences. 1998; 21(7):294–9. [PubMed: 9683321]

Cassidy BS, Leshikar ED, Shih JY, Aizenman A, Gutchess AH. Valence-based age differences in medial prefrontal activity during impression formation. Social Neuroscience. 2013; 8(5):462–473. http://doi.org/10.1080/17470919.2013.832373. [PubMed: 23998453]

Civettini A, Redlawsk D. Voters, emotions, and memory. Political Psychology. 2009; 30(1):125–151.

D'Argembeau A, Ruby P, Collette F, Degueldre C, Balteau E, Luxen A, Salmon E. Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. Journal of Cognitive Neuroscience. 2007; 19(6):935–44. http://doi.org/10.1162/jocn.2007.19.6.935. [PubMed: 17536964]

Denny, BT., Kober, H., Wager, TD., Ochsner, KN. A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex; Journal of Cognitive Neuroscience. 2012. p. 1-11.http://doi.org/10.1162/jocn_a_00233

Diano M, Celeghin A, Bagnis A, Tamietto M. Amygdala response to emotional stimuli without awareness: Facts and interpretations. Frontiers in Psychology. 2017 Jan.7:1–13. http://doi.org/10.3389/fpsyg.2016.02029.

Dolcos F, LaBar KS, Cabeza R. Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. NeuroImage. 2004a; 23(1):64–74. http://doi.org/10.1016/j.neuroimage.2004.05.015. [PubMed: 15325353]

Dolcos F, LaBar KS, Cabeza R. Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. Neuron. 2004b; 42(5):855–63. [PubMed: 15182723]

Dolcos F, LaBar KS, Cabeza R. Remembering one year later: role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. Proceedings of the National Academy of Sciences. 2005; 102(7):2626–31. http://doi.org/10.1073/pnas.0409848102.

Dougal S, Phelps EA, Davachi L. The role of medial temporal lobe in item recognition and source recollection of emotional stimuli. Cognitive, Affective & Behavioral Neuroscience. 2007; 7(3):233–42.

Gilron R, Gutchess AH. Remembering first impressions: Effects of intentionality and diagnosticity on subsequent memory. Cognitive, Affective & Behavioral Neuroscience. 2012; 12(1):85–98. http://doi.org/10.3758/s13415-011-0074-6.

Gozzi M, Zamboni G, Krueger F, Grafman J. Interest in politics modulates neural activity in the amygdala and ventral striatum. Human Brain Mapping. 2010; 31(11):1763–71. http://doi.org/10.1002/hbm.20976. [PubMed: 20162603]

Gutchess AH, Kensinger EA, Schacter DL. Functional neuroimaging of self-referential encoding with age. Neuropsychologia. 2010; 48(1):211–219. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2794905&tool=pmcentrez&rendertype=abstract. [PubMed: 19765600]

Hamann S. Cognitive and neural mechanisms of emotional memory. Trends in Cognitive Sciences. 2001; 5(9):394–400. [PubMed: 11520704]

Harris S, Kaplan JT, Curiel A, Bookheimer SY, Iacoboni M, Cohen MS. The Neural Correlates of Religious and Nonreligious Belief. PLoS ONE. 2009; 4(10):e7272. http://doi.org/10.1371/journal.pone.0007272.

Harvey P-O, Fossati P, Lepage M. Modulation of memory formation by stimulus content: specific role of the medial prefrontal cortex in the successful encoding of social pictures. Journal of Cognitive Neuroscience. 2007; 19(2):351–62. http://doi.org/10.1162/jocn.2007.19.2.351. [PubMed: 17280522]

Heatherton TF, Wyland CL, Macrae CN, Demos KE, Denny BT, Kelley WM. Medial prefrontal activity differentiates self from close others. Social Cognitive and Affective Neuroscience. 2006; 1(1):18–25. [PubMed: 18985097]

Kaplan JT, Freedman J, Iacoboni M. Us versus them: Political attitudes and party affiliation influence neural response to faces of presidential candidates. Neuropsychologia. 2007; 45(1):55–64. http://doi.org/10.1016/j.neuropsychologia.2006.04.024. [PubMed: 16764897]

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF. Finding the Self? An Event-Related fMRI Study. Journal of Cognitive Neuroscience. 2002; 14(5):785–794. http://doi.org/10.1162/08989290260138672. [PubMed: 12167262]

Kensinger EA, Schacter DL. Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. The Journal of Neuroscience. 2006; 26(9):2564–70. http://doi.org/10.1523/JNEUROSCI.5241-05.2006. [PubMed: 16510734]

Kim H. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. NeuroImage. 2011; 54(3):2446–61. http://doi.org/10.1016/j.neuroimage.2010.09.045. [PubMed: 20869446]

Knutson KM, Wood JN, Spampinato MV, Grafman J. Politics on the brain: An FMRI investigation. Social Neuroscience. 2006; 1(1):25–40. http://doi.org/10.1080/17470910600670603. [PubMed: 17372621]

Konkel A, Cohen NJ. Relational memory and the hippocampus: representations and methods. Frontiers in Neuroscience. 2009; 3(2):166–74. http://doi.org/10.3389/neuro.01.023.2009. [PubMed: 20011138]

Kumaran D, Maguire EA. Novelty signals: a window into hippocampal information processing. Trends in Cognitive Sciences. 2009; 13(2):47–54. http://doi.org/10.1016/j.tics.2008.11.004. [PubMed: 19135404]

LaBar KS, Cabeza R. Cognitive neuroscience of emotional memory. Nature Reviews Neuroscience. 2006; 7(1):54–64. http://doi.org/10.1038/nrn1825. [PubMed: 16371950]

Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Technical manual and affective ratings. NIMH Center for the Study of Emotion and Attention. 1997

Leshikar ED, Cassidy BS, Gutchess AH. Similarity to the self influences cortical recruitment during impression formation. Cognitive, Affective, & Behavioral Neuroscience. 2016; 16:302–314. http://doi.org/10.3758/s13415-015-0390-3.

Leshikar ED, Duarte A. Medial prefrontal cortex supports source memory accuracy for self-referenced items. Social Neuroscience. 2012; 7(2):126–45. http://doi.org/10.1080/17470919.2011.585242. [PubMed: 21936739]

Macrae CN, Moran JM, Heatherton TF, Banfield JF, Kelley WM. Medial prefrontal activity predicts memory for self. Cerebral Cortex. 2004; 14(6):647–54. http://doi.org/10.1093/cercor/bhh025. [PubMed: 15084488]

Mather M. Emotional arousal and memory binding: An object-based framework. Perspectives on Psychological Science. 2007; 2(1):33–52. [PubMed: 26151918]

Mitchell JP, Banaji MR, Macrae CN. The link between social cognition and self-referential thought in the medial prefrontal cortex. Journal of Cognitive Neuroscience. 2005; 17(8):1306–15. http://doi.org/10.1162/0898929055002418. [PubMed: 16197685]

Mitchell JP, Macrae CN, Banaji MR. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. The Journal of Neuroscience. 2004; 24(21):4912–7. http://doi.org/10.1523/JNEUROSCI.0481-04.2004. [PubMed: 15163682]

Mitchell JP, Macrae CN, Banaji MR. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron. 2006; 50(4):655–63. http://doi.org/10.1016/j.neuron.2006.03.040. [PubMed: 16701214]

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. Self-referential processing in our brain--a meta-analysis of imaging studies on the self. NeuroImage. 2006; 31(1):440–57. http://doi.org/10.1016/j.neuroimage.2005.12.002. [PubMed: 16466680]

Pais-Vieira C, Wing EA, Cabeza R. The influence of self-awareness on emotional memory formation: an fMRI study. Social Cognitive and Affective Neuroscience. 2016; 11(4):580–92. http://doi.org/10.1093/scan/nsv141. [PubMed: 26645274]

Paller KA, Wagner AD. Observing the transformation of experience into memory. Trends in Cognitive Sciences. 2002; 6(2):93–102. [PubMed: 15866193]

Phan KL, Wager T, Taylor SF, Liberzon I. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. NeuroImage. 2002; 16(2):331–48. http://doi.org/10.1006/nimg.2002.1087. [PubMed: 12030820]

Phelps, Ea, LeDoux, JE. Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron. 2005; 48(2):175–87. http://doi.org/10.1016/j.neuron.2005.09.025. [PubMed: 16242399]

Quamme JR, Yonelinas AP, Norman KA. Effect of unitization on associative recognition in amnesia. Hippocampus. 2007 Jan.200:192–200. http://doi.org/10.1002/hipo.

Ritchey M, LaBar KS, Cabeza R. Level of processing modulates the neural correlates of emotional memory formation. Journal of Cognitive Neuroscience. 2011; 23(4):757–71. http://doi.org/10.1162/jocn.2010.21487. [PubMed: 20350176]

Said CP, Baron SG, Todorov A. Nonlinear Amygdala Response to Face Trustworthiness: Contributions of High and Low Spatial Frequency Information. Journal of Cognitive Neuroscience. 2009; 21(3):519–528. http://doi.org/10.1162/jocn.2009.21041. [PubMed: 18564045]

Said CP, Dotsch R, Todorov A. The amygdala and FFA track both social and non-social face dimensions. Neuropsychologia. 2010; 48(12):3596–605. http://doi.org/10.1016/j.neuropsychologia.2010.08.009. [PubMed: 20727365]

Sakaki M, Niki K, Mather M. Beyond arousal and valence: The importance of the biological versus social relevance of emotional stimuli. Cognitive, Affective & Behavioral Neuroscience. 2012; 12(1):115–39. http://doi.org/10.3758/s13415-011-0062-x.

Santos A, Mier D, Kirsch P, Meyer-Lindenberg A. Evidence for a general face salience signal in human amygdala. NeuroImage. 2011; 54(4):3111–3116. http://doi.org/10.1016/j.neuroimage. 2010.11.024. [PubMed: 21081170]

Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, Parrish T. Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron. 2003; 39(4):701–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12925283. [PubMed: 12925283]

Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. Journal of Experimental Psychology: General. 1988; 117(1):34–50. [PubMed: 2966230]

Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. Neuropsychologia. 2009; 47(8–9):1765–79. http://doi.org/10.1016/j.neuropsychologia. 2009.02.028. [PubMed: 19428409]

Symons CS, Johnson BT. The self-reference effect in memory: A meta-analysis. Psychological Bulletin. 1997; 121(3):371–394. http://doi.org/10.1037/0033-2909.121.3.371. [PubMed: 9136641]

Talmi D, Anderson AK, Riggs L, Caplan JB, Moscovitch M. Immediate memory consequences of the effect of emotion on attention to pictures. Learning & Memory (Cold Spring Harbor, N.Y.). 2008; 15(3):172–82. http://doi.org/10.1101/lm.722908.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner Pa, Wood ER, Morris RGM. Schemas and memory consolidation. Science. 2007; 316(5821):76–82. http://doi.org/10.1126/science.1135935. [PubMed: 17412951]

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Morris RGM. Schema-dependent gene activation and memory encoding in neocortex. Science. 2011; 333(6044):891–5. http://doi.org/10.1126/science.1205274. [PubMed: 21737703]

Tsukiura T, Cabeza R. Remembering beauty: roles of orbitofrontal and hippocampal regions in successful memory encoding of attractive faces. NeuroImage. 2011; 54(1):653–60. http://doi.org/10.1016/j.neuroimage.2010.07.046. [PubMed: 20659568]

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002; 15(1):273–89. http://doi.org/10.1006/nimg.2001.0978. [PubMed: 11771995]

van Kesteren MTR, Beul SF, Takashima A, Henson RN, Ruiter DJ, Fernández G. Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. Neuropsychologia. 2013; 51(12):2352–9. http://doi.org/10.1016/j.neuropsychologia. 2013.05.027. [PubMed: 23770537]

van Kesteren MTR, Rijpkema M, Ruiter DJ, Fernández G. Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. The Journal of Neuroscience. 2010; 30(47):15888–94. http://doi.org/10.1523/JNEUROSCI. 2674-10.2010. [PubMed: 21106827]

van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN. How schema and novelty augment memory formation. Trends in Neurosciences. 2012; 35(4):211–9. http://doi.org/10.1016/j.tins.2012.02.001. [PubMed: 22398180]

Wagner DD, Haxby JV, Heatherton TF. The representation of self and person knowledge in the medial prefrontal cortex. WIREs Cognitive Science. 2012; 3(4):451–470. http://doi.org/10.1002/wcs.1183. [PubMed: 22712038]

Ward BD. Simultaneous inference for fMRI data. AFNI 3dDeconvolve Documentation, Medical College of Wisconsin. 2000

Winston JS, O'Doherty J, Kilner JM, Perrett DI, Dolan RJ. Brain systems for assessing facial attractiveness. Neuropsychologia. 2007; 45(1):195–206. http://doi.org/10.1016/j.neuropsychologia. 2006.05.009. [PubMed: 16828125]

Zald DH. The human amygdala and the emotional evaluation of sensory stimuli. Brain Research Reviews. 2003; 41(1):88–123. [PubMed: 12505650]

Zamboni G, Gozzi M, Krueger F, Duhamel J-R, Sirigu A, Grafman J. Individualism, conservatism, and radicalism as criteria for processing political beliefs: A parametric fMRI study. Social Neuroscience. 2009; 4(5):367–83. http://doi.org/10.1080/17470910902860308. [PubMed: 19562629]

Agreement with Beliefs Task (ABT)

Associative Recognition Task

Figure 1.

A. Arousal Ratings
(Normative study)

B. ABT Judgments
(In scanner)

C. Associative Recognition
(In scanner)
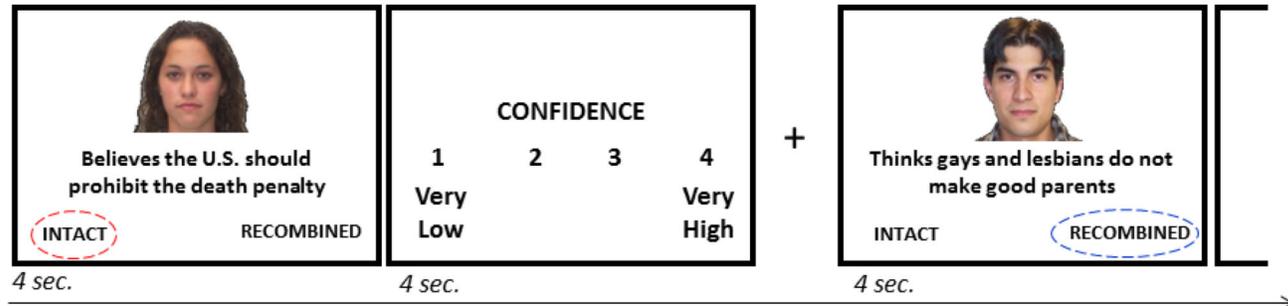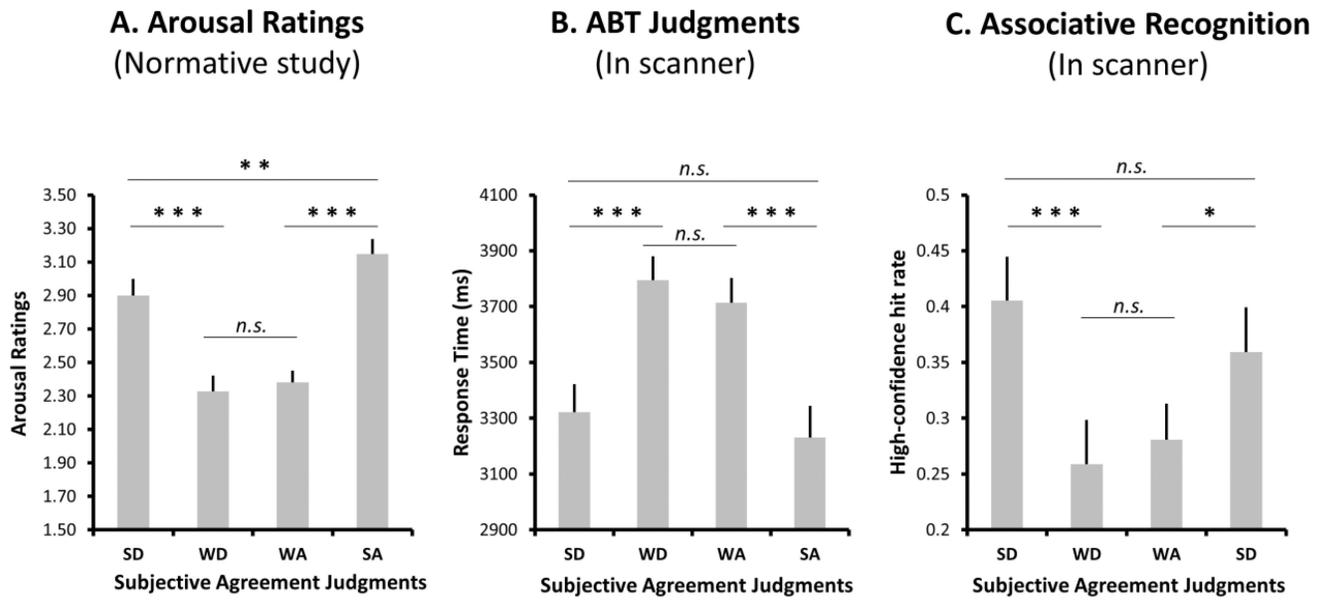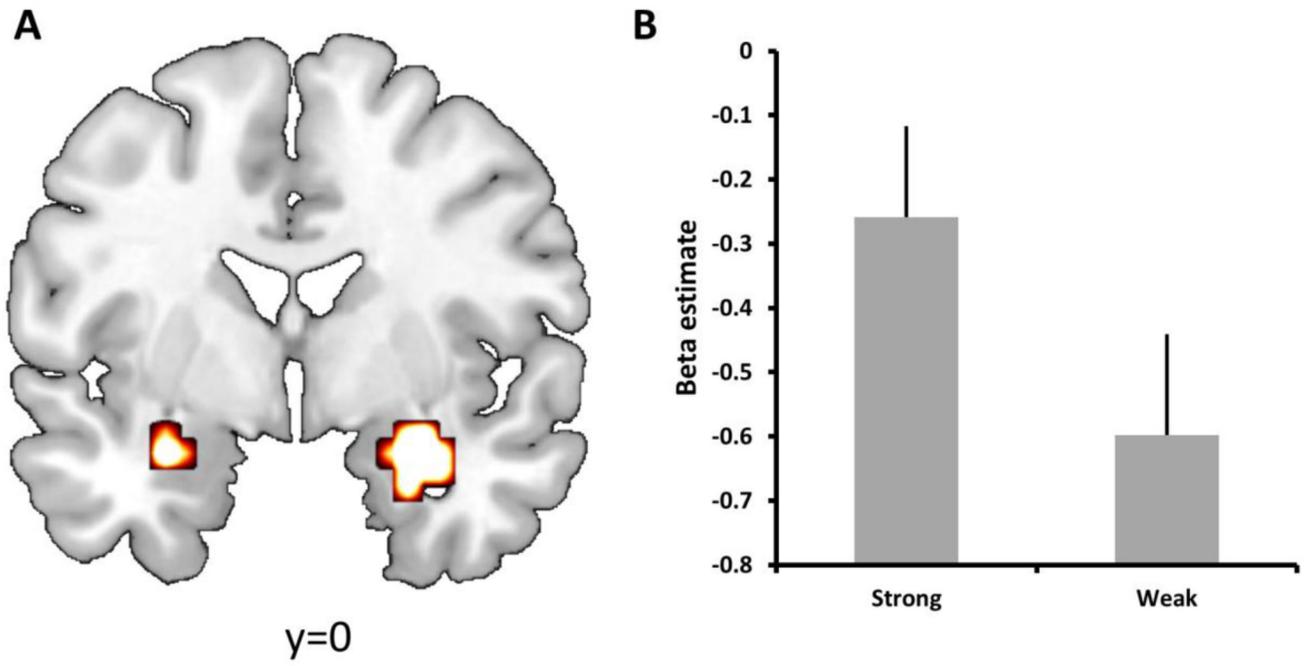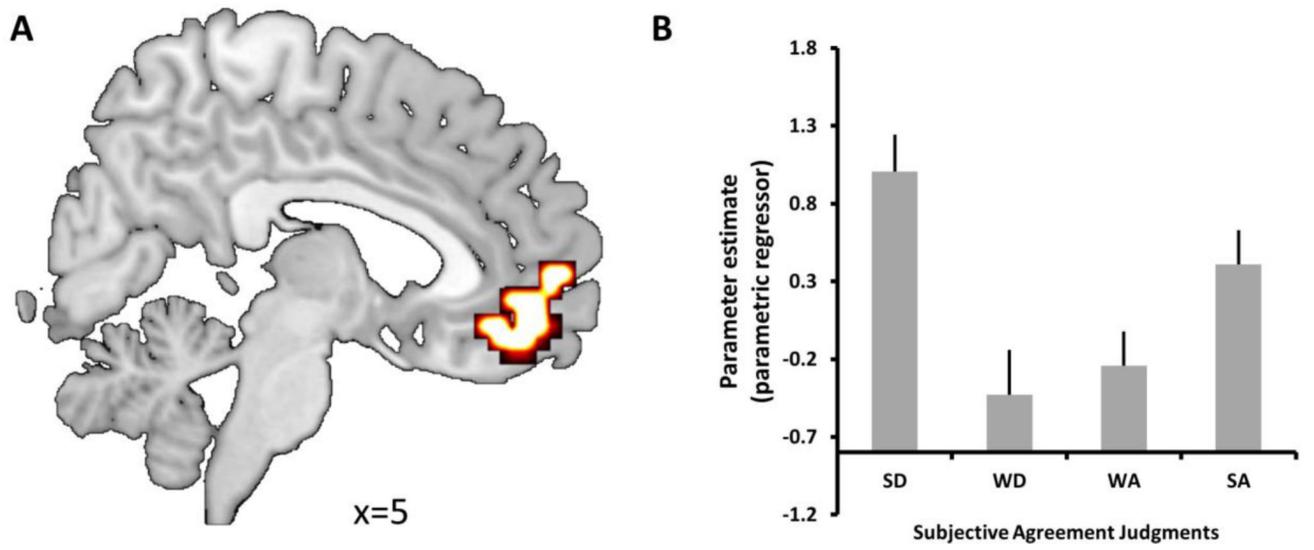
Figure 2.

Figure 3.

Figure 4.

Table 1

Main effects of agreement intensity and valence

Region	Hem	BA	MNI Coordinates			t	Voxels
			x	y	z		
Agreement Intensity (Strong/Weak)							
Strong > Weak							
Amygdala ROI	L	-	–30	–4	–19	3.74	4
Amygdala ROI	R	-	34	0	–19	5.50	18
Amygdala/rhinal cortex	R	-	34	–4	–19	5.59	44
Medial PFC/Anterior cingulate cortex	L	32	–8	45	15	5.40	290
Inferior frontal gyrus	R	45	49	38	15	4.60	30
Precentral gyrus	R	6	45	–8	49	4.43	37
Superior temporal gyrus	R	21	53	–56	23	4.85	37
Weak> Strong							
No significant clusters							
Agreement Valence (Agree/Disagree)							
Agree > Disagree							
Orbitofrontal cortex	L	11	–26	41	–8	5.16	100
Medial PFC	R	32	8	49	11	5.40	108
ACC/Superior frontal gyrus	L	32	–15	34	38	5.47	85
Precentral gyrus	R	3	34	–19	42	4.49	27
Cuneus	R	17	8	–83	4	6.54	66
Disagree > Agree							
No significant clusters							
Agreement Intensity × Valence Interaction							
Strong > Weak, Agree > Disagree							
Precentral gyrus	R	6	38	–15	49	5.61	39
Strong > Weak, Agree > Disagree							

Region	Hem	BA	MNI Coordinates			t	Voxels
			x	y	z		
No significant clusters							
Memory Modulated							
Agreement Intensity (Strong/Weak) × Subsequent Memory Interaction							
Medial PFC	-	11	−4	49	−11	5.52	29