**The Impact of Depression on Brain Activity During Source Memory Retrieval**

Elyssa M. Barrick and Daniel G. Dillon

Number of words:

Number of figures: 6

Number of tables: 2

Supplemental information: stimulus list, 1 table

Keywords:

Corresponding author:

Daniel G. Dillon, Ph.D.

Center for Depression, Anxiety and Stress Research

McLean Hospital

115 Mill Street, Belmont, MA 02478

Email: ddillon@mclean.harvard.edu

Phone: 617-855-4233

Fax: 617-855-4231

**Abstract**

**Background:** Recollection is disrupted in Major Depressive Disorder (MDD), but this disruption can be minimized by focused attention at encoding and retrieval. The neural mechanisms responsible for these clinically important phenomena are unclear. Thus, we used event-related potentials (ERPs) to examine recollection in MDD.

**Methods:** Twenty-four adults with MDD and 24 controls encoded words shown on the left or right (perceptual source) by making animacy or mobility judgments (conceptual source). ERPs were recorded during cued source retrieval, which depends on recollection.

**Results:** Mobility judgments prompted deeper encoding than animacy judgments, and memory was characterized by a *Group* x *Cue* x *Encoding Task* interaction: depressed adults were generally less accurate and less confident than controls, but they showed excellent conceptual source memory following deeper encoding. In parallel, a positive parietal ERP deflection that tracks recollection was globally reduced in depression, but sustained left parietal activation was seen during conceptual source judgments for deeply encoded words in depressed adults.

**Conclusions:** This study links two reliable effects of depression on recollection to electrophysiological activity over parietal cortex. First, accuracy and confidence were reduced in MDD, and the most reliable ERP correlate of recollection—a positive parietal deflection from 400-800 ms—was blunted. Second, depressed adults showed excellent memory when both the encoding and retrieval tasks demanded sustained attention, and this combination elicited sustained left parietal activity. These results link the impact of depression on recollection to left parietal activity, and they highlight the need for further work on this important topic.

**Introduction**

Memory retrieval plays a key role in Major Depressive Disorder (MDD) and, increasingly, in its treatment. Autobiographical memory retrieval in depression is “overgeneral” (Williams et al., 2007): cued to recall specific episodes from their lives, depressed adults tend instead to offer categorical accounts, summaries that average over distinct experiences and convey gist but few details. This lack of precision has consequences—overgeneral autobiographical retrieval predicts a longer course of depression (Brittlebank, Scott, Williams, & Ferrier, 1993; Peeters, Wessel, Merckelbach, & Boon-Vermeeren, 2002; Sumner, Griffith, & Mineka, 2010), perhaps because learning to cope with MDD depends on accurately remembering (and learning from) past struggles. Moreover, enhancing retrieval specificity has emerged as a promising treatment for depression. A pilot study with inpatients found that practice retrieving specific, detailed memories reduced hopelessness (a potent risk factor for suicidality; Aaron T Beck, Steer, Kovacs, & Garrison, 1985) and brooding rumination while enhancing problem solving (Raes, Williams, & Hermans, 2009). Another study used the same approach to treat depression in Afghan refugees, adding a wait-list control group (Neshat-Doost et al., 2012). Memory training acutely reduced depressive symptoms and led to improvements at two months’ follow-up that were stronger in the treated group versus the controls. In short, memory retrieval is impaired in depression, this disruption is clinically important, and targeting retrieval for intervention can bring lasting relief.

Given these facts, the depth of our ignorance regarding the neurobiology of memory retrieval in depression is astonishing, particularly since the cognitive neuroscience of episodic retrieval in healthy adults has been studied extensively (e.g., Eichenbaum, Yonelinas, & Ranganath, 2007; Rugg & Curran, 2007; Rugg & Vilberg, 2013). The does not reflect lack of desire: a decade ago, the National Institutes of Mental Health, Aging, and Neurological Disorders and Stroke called for integration of research on depression and memory (Steffens et al., 2006). Furthermore, it is not as though the nature of the deficit is unclear. As one might expect based on the overgeneral memory literature, several studies report that depression leaves familiarity (or “automatic”) memory intact but impairs recollection—the retrieval of contextual details that specify the spatiotemporal source of a memory (G M MacQueen, Galway, Hay, Young, & Joffe, 2002; Glenda M MacQueen et al., 2003; Raes et al., 2006; Ramponi, Barnard, & Nimmo‐Smith, 2004). The source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993) provides a detailed account of the cognitive processes that mediate recollection, and this has supported dozens of event-related potential (ERP) and functional magnetic resonance imaging (fMRI) studies of source memory retrieval in healthy adults. Unfortunately, a parallel literature on disrupted source memory retrieval in depression has not emerged.

This study takes a step towards addressing this glaring gap in the literature by using ERPs to study source memory retrieval in unmedicated adults with MDD. We used neutral stimuli to avoid confounds associated with mood congruent encoding (G H Bower, 1981; Gordon H. Bower, 1987; Dillon, Dobbins, & Pizzagalli, 2014), adapting a design that can tease apart neural systems engaged by retrieval of conceptual versus perceptual aspects of memory (Bergström, Henson, Taylor, & Simons, 2013; Dobbins & Wagner, 2005; Simons, Gilbert, et al., 2005). At study, participants viewed words presented on the left or right side of a computer monitor above a question that specified one of two encoding tasks—an animacy judgment or a mobility judgment. At test, the participants were cued to retrieve the presentation side (perceptual source) and the encoding task (conceptual source) on separate trials.

A recent study (Bergström et al., 2013) used fMRI to show that both conceptual and perceptual retrieval recruited the precuneus, a structure in the posterior midline, and electrophysiological data collected in the same participants related this fMRI result to a sustained, negative polarity ERP maximal over posterior electrodes and commonly referred to as the late posterior negativity, or LPN (Cycowicz, Friedman, & Snodgrass, 2001; Johansson & Mecklinger, 2003; Mecklinger, Johansson, Parra, & Hanslmayr, 2007). Critically, the study found that the LPN extended over left frontal cortex only during conceptual retrieval, and this was mirrored by dorsolateral PFC activation in the fMRI session. Finally, both conceptual and perceptual retrieval elicited the most well-studied ERP marker of recollection—a positive deflection over parietal sites that extends from about 400-800 ms post-stimulus, often with a left hemisphere maximum (Rugg & Curran, 2007). The timing and distribution of these ERPs suggests that retrieval attempts first activate parieto-hippocampal circuits, bringing candidate memories to mind and generating the parietal ERP effect from 400-800 ms, and then the candidate memories are inspected and elaborated upon until one is selected. When the retrieval attempt is directed at perceptual information, the second stage engages posterior cortical regions, but when conceptual information is sought, activation of left PFC regions known to support semantic encoding, elaboration, and selection among alternative representations is seen (Badre & Wagner, 2007). Because MDD is associated with volumetric loses in hippocampus and PFC (e.g., Treadway et al., 2015), and because rumination may occupy left PFC circuits that would be engaged by conceptual source retrieval, we expected MDD to particularly impair conceptual source memory.

However, in the course of our analysis it became obvious that we had overlooked a critical factor. Over three decades, Hertel has amassed a wealth of data showing that although recollection is typically impaired by depression, poor performance is not inevitable (P. T. Hertel & Brozovich, 2010; P.T. Hertel & Rude, 1991; Paula T. Hertel, 1997; Paula T. Hertel, Benbow, & Geraerts, 2012; Paula T. Hertel & Hardin, 1990). If attention is sustained at encoding, retrieval, or both, depressed adults can show excellent memory, in some cases outperforming controls (P.T. Hertel & Rude, 1991). As will be seen below, one of our tasks promoted deeper encoding than the other, and when words from that task were targeted for conceptual source retrieval, the MDD group performed quite well. Thus, this study highlights neural mechanisms linked to disrupted source memory in MDD, but also brain activity that supports excellent memory when encoding and retrieval conditions are salubrious.

**Materials and Methods**

**Participants**

Participants (18-62 years old, right-handed, with no history of neurological or unstable medical conditions) were recruited from the community and compensated ($25.00/hour) for their time. Consent was obtained with a protocol approved by the Partners HealthCare Human Research Committee. To determine eligibility, prospective participants were first screened by phone or online. Controls had to report no current or past psychiatric conditions or unstable medical illness. Participants in the MDD group had to report current depression, no history of other DSM-IV Axis I diagnosis (excepting secondary generalized anxiety, social anxiety, and specific phobia due to their high comorbidity with MDD), and no medication use in the past 2 weeks (6 weeks for fluoxetine, 6 months for neuroleptics). Thirty-four controls and 30 depressed adults were invited to complete the ERP session.

To confirm that the screening was accurate, immediately after each ERP session we assessed psychiatric history with the MINI International Neuropsychiatric Interview, version 6.0 (Sheehan et al., 1998) and administered the Beck Depression Inventory II (BDI-II; A.T. Beck, Steer, & Brown, 1996). Data from depressed participants (*n* = 26) were retained provided they met criteria for MDD but no other DSM-IV Axis I diagnosis with previously noted exceptions, and had a BDI-II score ≥ 14. Data from 2 depressed and10 healthy individuals were ultimately excluded due to EEG artifacts (see section *EEG pre-processing*). Thus, the final sample included 24 unmedicated adults with MDD and 24 healthy controls.

**Self-report Measures**

In addition to the BDI-II, we administered the Mood and Anxiety Symptom Questionnaire(MASQ; Watson et al., 1995), the Ruminative Responses Scale (RRS; Treynor, Gonzalez, & Nolen-hoeksema, 2003), and the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) (PSQI). The Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) was used as a brief IQ assessment. We included these measures to characterize the MDD sample and to test whether depression-related deficits were related to variation in depressive severity (BDI-II total), brooding rumination (RRS-Brooding total), and/or sleep disturbance (PSQI total). One depressed participant did not complete the PSQI.

**Task**

The task was programmed in PsychoPy (Peirce, 2008). Due to a hardware change, RT data were not acquired for one control and one depressed participant.

**Stimuli.** Participants made animacy and mobility judgments for words over 100 encoding trials. Therefore, we used the MRC Psycholinguistic Database (Coltheart, 1981) to select 25 words from four categories: “living/immobile” (e.g., *oak*), “non-living/immobile” (e.g., *shed*), “living/mobile” (e.g., *dog*), and “non-living/mobile” (e.g., *kite*). One-way analyses of variance (ANOVA) yielded no significant differences among the lists for number of letters (mean±S.D. = 5.27±1.29), number of syllables (1.52±0.50), frequency of occurrence (35.58±79.02), concreteness (598.87±20.18), or imageability (596.80±25.31), all *ps* > 0.064. Words are listed in the Supplement.

**Encoding.** The task included six encoding-retrieval cycles. Each encoding block included 16 trials (Figure 1, *left*) in which a word appeared on the left or right side directly above one of two questions: “living/non-living?” (animacy judgment) or “mobile/immobile?” (mobility judgment) (duration: 3.5 s). Participants responded by pressing a button. Thus, each word was encoded in relationship to a perceptual source defined by screen position (left vs. right) and a conceptual source defined by the encoding task (animacy vs. mobility). A jittered inter-trial interval (500-2000 ms) separated the trials.

Immediately after encoding, a 3-digit number (262, 931, 888, 704, 557, or 474) was presented for 30 s and participants counted backwards from that number, in steps of three, out loud. This task was used to disrupt rehearsal and clear working memory (Reitman, Higman, Lifson, & Rosenblum, 1974). The experimenter observed the counting but no data were collected.

PLEASE INSERT FIGURE 1 ABOUT HERE

**Retrieval.** Each retrieval block comprised 48 trials that included a cue, word, and response screen (Figure 1, *right*). On 16 trials each, the cue was “Side” or “Question” and the word came from the immediately preceding encoding block; these cues prompted retrieval of the perceptual and conceptual sources, respectively. On the remaining 16 trials, the cue was “Odd/Even” the word was a numeral between “one” and “ninety-six”, and the participant was asked to judge parity. Odd/Even trials served as a control condition: participants had to read a cue, interpret it, and retrieve information before responding, but—in contrast to Side and Question trials—retrieval was directed towards semantic rather than episodic memory. Thus, comparing ERP data from Side or Question trials versus Odd/Even trials should isolate activity mediating episodic retrieval. Presentation order of words and cues was randomized.

The response screen consisted of ‘RESPOND’ printed above the word with the numbers 1-5 printed below and corresponding to a choice and the level of confidence in that choice: 1 = high confidence (*left* or *living/non-living* or *odd*), 2 = low confidence (*left* or *living/non-living* or *odd*), 3 = guess, 4 = low confidence (*right* or *mobile/immobile* or *even*), 5 = high confidence (*right* or *mobile/immobile* or *even*). Participants were told to select “guess” when they were unable to retrieve any information from encoding. Finally, a centrally presented fixation cross was visible throughout retrieval, and a jittered inter-trial interval (500-2000 ms) separated the trials. Similar methods have been used in previous ERP and fMRI studies of cognitive versus perceptual source retrieval in healthy adults (Bergström et al., 2013; Simons, Gilbert, et al., 2005; Simons, Owen, Fletcher, & Burgess, 2005).

**Procedure.** Following application of the EEG net, participants received detailed instructions. For the mobility task, participants were told to respond “mobile” if the word described an object whose primary purpose is to move (e.g., *car*) or that can move under its own power (e.g., *dog*); otherwise, they were told to respond “immobile”. Participants then completed a practice cycle with ten retrieval trials (four Side, four Question, and two Odd/Even).

**EEG Recording**

The EEG was recorded during retrieval with a 128-sensor HydroCel GSN Electrical Geodesics Inc (EGI) net connected to a Net Amps 300 amplifier (sample rate = 1000 Hz, 0.02–100 Hz bandpass filter). Data were referenced to the vertex during acquisition. Impedances were kept below 45 kΩ whenever possible; none exceeded 75 kΩ.

**Behavioral Data Analysis**

We analyzed the behavioral data using linear mixed models implemented with the R (R Developement Core Team, 2015) library *lme4* (Bates, Maechler, Bolker, & Walker, 2015). This method is preferable to ANOVA because it is easy to include multiple covariates that might influence source memory (e.g., age: Mark & Rugg, 1998) or depression (e.g., gender: Nolen-Hoeksema, 2001), thus improving the likelihood of isolating the unique effect of depression on source memory. Moreover, there is no need to aggregate the data into subject-level means, thus improving estimates of within and between-subject variability. The specific models for each dependent variable are described below, but in all cases we computed a first model with task elements (e.g., encoding tasks, retrieval cues) and covariates (age, gender) as fixed effects but with *Group* omitted. Next, we added *Group* in a second model, either as a main effect or in interaction with other factors, and then used likelihood ratio tests implemented in the R library *anova* to compare model fits by chi-square. If the second model was a significant improvement, we report its parameters; otherwise, we report parameters from the first model. All models used *Word* and *Subject* as random effects to account for individual differences not attributable to group and variability in the memorability of particular words. When modeling encoding accuracy (coded 0 or 1), we used glmer with the logit link function. Finally, we extracted *p*-values via the R library *lmerTest*.

**Encoding**. We dropped trials (< 1%) with no response or where RT exceeded the participant’s mean±3SD. Our first accuracy model included *Task* (animacy, mobility), *Side* (left, right), *Block*, *Gender*, and *Age* as fixed effects. Our first RT model included the same factors plus *Accuracy*. We compared these to models that included *Group*.

**Retrieval**. We dropped trials with no response or where RT exceeded the participant’s mean±3SD (< 2%). Next we analyzed the Odd/Even trials to judge their suitability as a control condition.Accuracy (percent correct) on these trials was at ceiling (controls: 98.43±0.12; MDD: 99.13±0.09), and RT (in ms) was similar between groups (controls: 862.58±51; MDD: 779.00±48). Adding *Group* did not improve models that included *Block, Age*, and *Gender*, χ2s < 2.1, *p*s > 0.14. Thus, the Odd/Even trials elicited similar behavior across groups and are therefore suitable as a control condition.

For the Side and Question trials, accuracy was coded as follows: incorrect, high confidence = 1; incorrect, low confidence = 2; guess = 3; correct, low confidence = 4; correct, high confidence = 5. Confidence was coded guess = 1, low confidence = 2, high confidence = 3. We computed three models for accuracy and confidence. The first included *Block*, *Cue* (Side, Question), *Encoding Task* (animacy, mobility), *Encoding Side* (left, right), *Age*, and *Gender*. The second model added a *Cue* x *Encoding Task* interaction, and the third added a *Group* x *Cue* x *Encoding Task* interaction, which automatically added the main effect and all two-way interactions involving *Group*. We used similar models to analyze RT on correct trials.

**ERP Analysis**

**Pre-processing.** Pre-processing was conducted with the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes for MATLAB (MathWorks, Natick). EEG data were merged across blocks and re-referenced to the average of all electrodes prior to bandpass filtering (0.1-30 Hz). Bad channels were visually identified and interpolated. We set an *a priori* cut-off of up to 18 bad channels (14% of the total); data from three controls exceeded this cut-off and were excluded. Next, data marked by gross artifacts were manually removed and independent component analysis was used to extract components capturing blinks, HEOG, and EKG. The cleaned data were time-locked to word onsets and segmented (-200 to 2000 ms), and the pre-stimulus interval was used for baseline correction. We rejected segments where the maximum-minimum voltage difference (computed over 200 ms intervals, 100 ms sliding window) exceeded 100 μV, or where any raw value exceeded +/- 100 μV.We removed data from seven controls and two depressed adults where over 50% of trials were rejected, yielding the final sample (*n* = 24 per group). The mean number of clean segments in each bin defined by *Group* x *Cue* x *Task* ranged from 21-28 for source hits.Guesses were excluded from the ERP analysis and there were too few segments per bin for analysis of source misses (range of means: 6-11). Finally, the data were averaged for group-level analyses.

**Group-level analyses**. Because we initially expected a negative effect of MDD on conceptual source memory regardless of the encoding tasks, we first inspected the ERP waveforms associated with all correct responses to the Question, Side, and Odd/Even cues, collapsed over encoding task. There was an obvious difference in the left parietal ERP from 400-800 ms (see Results), thus we extracted mean amplitude data over this time window from parietal electrodes in the left (P1, P3, P5, P7) and right (P2, P4, P6, P8) hemispheres and submitted them to a *Group* x *Cue* x *Hemisphere* ANOVA. This method has been used in many prior studies (Hayama, Johnson, & Rugg, 2008; Rugg & Curran, 2007; Wilding & Rugg, 1996), and the focus on a few electrodes, over one time window, constitutes a traditional ERP analysis.

Further analyses used a newer approach and were intended to parallel the memory accuracy data, which yielded a *Group* x *Cue* x *Encoding Task* interaction. To identify brain activity that might mediate the interaction, we computed Question minus Side difference waves separately for words from the mobility and animacy encoding tasks in each group. We then submitted the difference waves to mass univariate analysis in order to identify significant activations within and across groups (Groppe, Urbach, & Kutas, 2011a), focusing on mean amplitudes from 400-800 ms, 800-1400 ms, and 1400-2000 ms post-stimulus. In contrast to the traditional approach to ERPs, mass univariate analysis is similar to methods commonly used in fMRI research (Friston et al., 1995) and involves computing a either a one-sample *t*-test against zero (within-group analysis) or a two-sample *t*-test (between-group analysis) at each electrode. This makes much better use of the spatiotemporal richness of ERP data than the traditional method. However, appropriate correction for multiple comparisons can degrade power. To limit this concern, we used a cluster-based permutation procedure that can detect moderately strong effects provided they extend over several electrodes (Groppe, Urbach, & Kutas, 2011b). All electrodes within 4 cm of each other were considered neighbors, and neighboring electrodes with *p*-values less than 0.05 (uncorrected) were considered clusters. The sum of all *p­*-values within a cluster constituted its mass. We then performed 2500 permutations (within or across groups, depending on the test), selecting the most extreme cluster mass score from each permutation in order to generate a sampling distribution of cluster mass values (Bullmore et al., 1999). This distribution was used to judge the probability of observing clusters of the size seen in the current dataset; only clusters significant at *p* < 0.05 were considered reliable. This approach maximizes power while maintaining a familywise error rate of approximately 5%.

**Results**

**Demographics**

There were no group differences in gender, age, or education (Table 1). As expected, the MDD group endorsed more depression and anxiety than the controls, with the mean BDI-II score indicating moderate depression. The MDD group also reported more rumination and poorer sleep, but there was no difference in IQ as estimated from WTAR scores.

PLEASE INSERT TABLE 1 ABOUT HERE

**Behavior**

**Encoding**. Encoding was not affected by depression but the mobility task was harder than the animacy task, judging by percent correct (mobility: 92.42±0.26; animacy: 95.85±0.20; *Z* = -4.91, *p* < 0.001) and RT (mobility = 1,801±552 ms; animacy = 1,664±535 ms; *Z* = 10.54, *p* < 0.001). Participants responded faster when making correct (1,720±541 ms) vs. incorrect (1,923±619 ms) judgments, *Z* = -3.46, *p* < 0.001, and RT decreased over the session (linear *Block* effect, *Z* = -6.34, *p* < 0.001). Adding *Group* did not improve the accuracy or RT models, χ2s < 1.93, *p*s > 0.16, thus depressed and healthy adults performed similarly.

**Source accuracy**. In contrast to encoding, source accuracy was influenced by depression and the addition of *Group* improved the model, χ2 = 26.40, *p* < 0.001. There was a *Group* x *Cue* x *Encoding Task* interaction, *Z* = -2.13, *p* = 0.033, which subsumed significant *Cue* x *Encoding Task* and *Group* x *Encoding Task* interactions and main effects of *Cue* and *Encoding Task* (*Z*s > 2.7, *p*s < 0.006). Figure 2A (left panel) shows that the triple interaction emerged because accuracy under the Side cue did not vary substantially by encoding task, but accuracy under the Question cue was better following mobility versus animacy judgments, and this effect was larger in MDD. As shown in Figure 2A (right panel), we quantified this by computing Question minus Side accuracy difference scores for each encoding task. In the MDD group, the mean score was positive for the mobility task but negative for the animacy task, and this difference was significant, *t*(23) = 4.47, *p* < 0.001, *d* = 0.91. In controls the mean score was negative for both tasks but was marginally more positive for the mobility task, *t*(23) = 2.04, *p* = 0.053, *d* = 0.42. A between-groups test on difference scores for the animacy task was not significant, *t*(46) = 0.77, *p* = 0.44, *d* = 0.22. However, the difference score for the mobility task was more positive in depressed adults, *t*(46) = 3.04, *p* = 0.004, *d* = 0.88.

Because the ERP analysis focused on hits, we repeated this analysis with hit rates (percent correct), which are given in Table 2. A *Group* x *Cue* x *Encoding Task* ANOVA on these data revealed effects of *Cue*, *F*(1, 46) = 17.47, *p* = 0.001 and *Encoding Task*, *F*(1, 46) = 31.95, *p* < 0.001, as well as two significant interactions: *Group* x *Cue*, *F*(1, 46) = 6.42, *p* = 0.015, and *Cue* x *Encoding Task*, *F*(1, 46) = 22.35, *p* < 0.001. The *Group* x *Cue* x *Encoding Task* interaction was not significant, *F* < 1. However, analysis of Question minus Side accuracy difference scores yielded essentially the same results as described above. For depressed adults the Question minus Side accuracy difference score was more positive for words from the mobility task (5.03±11.95) versus the animacy task (-8.76±9.82), *t*(23) = 3.82, *p* = 0.001, *d* = 1.26. The same was true for controls (mobility: -2.97±9.74; animacy: -12.27±14.11; *t*(23) = 2.83, *p* = 0.010, *d* = 0.76). Once again, there was no group effect for the animacy task, *t*(46) = 1.00, *p* = 0.32, *d* = 0.29, but the Question minus Side difference score was significantly larger in depressed versus healthy adults for the mobility task, *t*(46) = 2.54, *p* = 0.015, *d* = 0.73.

Thus, depressed adults were more accurate than controls when responding to words from the mobility task presented under the Question cue. Indeed, simple pairwise comparisons showed better accuracy for depressed versus healthy adults in the Question/mobility cell of the design, *Z* = 1.98, *p* = 0.048. However, this was the only cell in which the MDD group outperformed the controls—in the other three cells accuracy was non-significantly higher in healthy adults (*Z*s < 1.63, *p*s > 0.10). In other words, depressed adults generally performed worse than controls, except when words from the mobility task were presented under the Question cue. Finally, the best-fitting model included effects of *Age*, *Z* = -3.31, *p* < 0.002, *Gender*, *Z* = 3.32, *p* < 0.002, and *Block, Z* = 3.23, *p* = 0.001, reflecting higher accuracy in younger vs. older adults, in men vs. women, and in later vs. earlier blocks.

PLEASE INSERT FIGURE 2 AND TABLE 2 ABOUT HERE

**Source confidence**. Figure 2B (top panel) shows that depressed adults were less confident than controls. Consistent with this impression, the model was improved by the addition of *Group*, χ2 = 18.46, *p* = 0.001, and included a trend for a *Group* x *Cue* interaction, *Z* = -1.65, *p* = 0.098, as the difference in confidence was stronger under the Side cue, *Z* = 2.42, *p* = 0.016, than the Question cue, *Z* = 1.14, *p* = 0.255. The model also revealed main effects of *Cue*, *Z* = -5.33, *p* < 0.001 and *Encoding Task*, *Z* = 2.91, *p* = 0.004, as well as a linear effect of *Block*, *Z* = 3.69, *p* < 0.001; participants were more confident under the Question cue, when responding to words from the mobility task, and when responding in later blocks.

**Source RT**. The bottom panel of Figure 2B shows that correct RT was similar across the groups, with all participants slower in response to the Question vs. Side cue. Accordingly, the first model was not improved by the addition of *Group*, *p =* 0.08, but it included a strong effect of *Cue*, *Z* = -45.51, *p* < 0.001. There was also a negative linear effect of *Run*, *Z* = -18.82, *p* < 0.001, a strong effect of *Confidence*, *Z* = 21.61, *p* < 0.001, and an effect of *Gender*, *Z* = -3.09, *p* = 0.003. These reflected shorter correct RTs in later vs. earlier blocks, for high vs. low confidence responses, and in males vs. females.

**ERPs**

Figures 3 displays waveforms elicited by correct responses to the Question, Side, and Odd/Even cues. In controls, the three ERPs most frequently associated with episodic retrieval are apparent. From 400-800 ms there is robust activity over parietal electrodes. This ERP has been consistently associated with recollection (Rugg & Curran, 2007), and it is noticeably stronger on Question and Side trials relative to Odd/Even trials. From 800-2000 ms, a positive potential is evident over right frontal cortex. This potential is elicited during decision-making tasks and is thought to reflect monitoring of the product of each retrieval attempt in order to render a response (Fleck, Daselaar, Dobbins, & Cabeza, 2006; Hayama et al., 2008; Hayama & Rugg, 2009); in contrast to the parietal deflection, it is also evident during Odd/Even judgments. Finally, the LPN is apparent during the same time window (800-2000 ms) as the right frontal effect (Cycowicz et al., 2001). The LPN is not elicited on Odd/Even trials and has a medial posterior focus, but it extends into left frontal cortex on Question trials.

The results from depressed adults are similar but parietal activity was markedly weaker. As outlined in the Methods, we extracted mean amplitudes between 400-800 ms from parietal electrodes and conducted a *Group* x *Cue* x *Hemisphere* ANOVA. This yielded a main effect of *Group*, *F*(1,46) = 4.35, *p* = 0.043, due to decreased amplitude in MDD. There was also a main effect of *Cue*, and follow-up tests yielded reliable differences between all conditions (REGWQ; Question > Side > Odd/Even, *ps* < 0.043). These data suggest that recollection was strongest under the Question cue and reduced in MDD.

PLEASE INSERT FIGURE 3 ABOUT HERE

To test our *a priori* hypothesis, we subtracted activity on Odd/Even trials from Question and Side trials. Figure 4 shows that in both groups, the Question condition drove activity over parietal electrodes from 400-800 ms with a left hemisphere focus, although this effect was only reliable in controls (a negative difference over right frontal cortex in this window was only significant in the MDD group). From 1400-2000 ms, this subtraction revealed a negative difference over left PFC in both groups. By contrast, the Side condition elicited an LPN over midline posterior sites in both groups from 800-1400 ms. Contrary to our prediction, between-groups tests on these difference waves yielded no reliable findings.

PLEASE INSERT FIGURE 4 ABOUT HERE

***Group x Cue x Encoding Task***. Our planned ERP analysis did not reveal group differences, but it was not designed to capture the *Group* x *Cue* x *Encoding Task* that characterized source accuracy. To address this limitation, we computed Question minus Side difference waves separately for words from each encoding task and then submitted them to within and between-groups analysis, duplicating our approach to the accuracy data. Figure 5 shows the results; see Table 3 for details. In depressed adults, the Question minus Side difference varied drastically by encoding task. Words from the mobility task elicited sustained activation over left parietal cortex, leading to significant effects in all three time windows. By contrast, words from the animacy task elicited a negativity over fronto-central sites that was stronger over the left hemisphere; again, significant differences were observed in all time windows. In controls, no significant differences were observed for either task in any time window. Finally, between-groups comparisons revealed no differences for the animacy task, but there were reliable differences for the mobility task over left centro-parietal electrodes from 400-800 and 800-1400 ms, with stronger activation in the MDD group. Thus, the Question (minus Side) cue had a strong, task-dependent effect on brain activity in MDD that was muted in controls.

PLEASE INSERT FIGURE 5 AND TABLE 3 ABOUT HERE

**Correlational Analyses**

We computed Pearson correlations in the MDD group to determine if individual differences in depressive severity, brooding rumination, or sleep disruption were related to source accuracy or to ERPs sensitive to MDD. We found no relationship with source accuracy, with left parietal activity averaged over encoding tasks, or with left frontal activity on Question/animacy trials (|*p*s| < 0.31, *p*s > 0.15). However, the magnitude of parietal activity isolated by the Question minus Side subtraction for words from the mobility task (Figure 5, *first row*) was affected by sleep disruption. Specifically, as shown in Figure 6, ERP amplitudes from 400-800 ms and 800-1400 ms were negatively related to PSQI total scores (*r*s < -0.47, *ps* < 0.02).

PLEASE INSERT FIGURE 6 ABOUT HERE

**Discussion**

This study yielded two sets of parallel behavioral and ERP results. First, relative to controls the depressed adults were generally less accurate and less confident in their source memories, and this was mirrored by a significant reduction in parietal ERP amplitude from 400-800 ms. The negative effect of depression on memory was relatively modest, but it should be noted that in addition to generating lower confidence ratings than controls in all four cells of the design, the depressed adults generated lower mean accuracy scores than controls in three cells of the design; worse performance by one group in 7/8 cells analyzed is improbable under the null-hypothesis (binomial test, *p* = 0.035 one-tailed). Second, depressed adults showed excellent memory for words from the mobility task presented under the Question cue, and this was mirrored by robust activity over left parietal cortex that exceeded what was seen in controls. Consequently, depressed adults appear to have under-recruited parietal circuits on the bulk of trials, leading to source memory impairment, but then strongly activated these same circuits on Question/mobility trials, leading to a sharp improvement in memory accuracy.

The first set of findings conforms to our expectations, which were informed by the prior literature. It is clear from behavioral research that depressed adults show recollection deficits, and thus we expected them to perform worse than controls on this challenging source memory task. Because parietal activity from ~400-800 ms is a marker of recollection (Curran & Rugg, 2007), we expected depression to reduce the amplitude of this ERP, as it did. However, the second set of findings was unexpected—but perhaps it should have been anticipated. Hertel has consistently found negative effects of depression on memory when encoding tasks are loosely structured and leave free time for mind-wandering, and she has linked this to off-task thoughts in depression (REF). By contrast, when encoding tasks demand sustained attention and thus admit little mind-wandering, she has found excellent memory in depression (Hertel & Rude, 1991). Both of our encoding tasks offered structure, and both are conventionally considered “deep” tasks. However, both groups took longer to complete the mobility task versus the animacy task, and this was accompanied by a difference in accuracy—probably because judging animacy is straightforward, but judging mobility is difficult for some words (e.g., *oak* would be considered immobile in the context of instructions provided to participants, but oak trees sway in the breeze). Therefore, we argue that the mobility task required more sustained attention and prompted deeper encoding than the judgment task, thus offering richer targets for retrieval under the Question cue. Moreover, responses to the Question cue were slower than responses to the Side cue, indicating that participants labored over conceptual retrieval more than perceptual retrieval. Consequently, Question/mobility trials demand sustained attention at encoding and retrieval, and Hertel’s work predicts that this is when depressed adults should do best—as indeed they do.

Until now it has not been possible to do more than speculate about the brain systems that support good source memory in depression when this is observed, because (to our knowledge) there is only one prior functional study of source memory in depression, and that study found no group differences in behavior (REF). Here we show that the same activity implicated in source memory failure is implicated in source memory success following deep encoding—namely, left parietal cortex. Correct responses on Question/mobility trials elicited strong activity over left parietal cortex extending over all time windows analyzed in depressed adults, and this activation exceeded that seen in controls from 400-800 and 800-1400 ms. Because of the aforementioned link to recollection, these ERP effects suggest that the combination of deep encoding and conceptual retrieval supports the sustained retrieval of rich detail in depressed adults. By contrast, when conceptual retrieval was directed at words from the shallower encoding condition (animacy task), the depressed adults generated a lasting negativity over left PFC. As outlined in the Introduction, this likely reflects some combination of cue elaboration or selection amongst competing memory representations, which are probably more necessary in the context of this more difficult cell of the design.

These results provide valuable initial insights into source memory retrieval in depression, but some limitations should be mentioned. First, the negative effect of MDD on memory was relatively modest, and it may be useful to modify the design and/or the recruitment strategy in order to identify more severe disruptions. In fact, our use of unmedicated outpatients and neutral stimuli, in the context of a fully anticipated memory test with no stress manipulation, is probably the

**Figure Captions**

*Figure 1*. Encoding (left) and recognition (right) trial structures. Encoding trials began with three centrally presented arrows pointing to the side on which the word would appear. The encoding task was presented next, either “living or non-living?” (animacy judgment) or “mobile or immobile?” (mobility judgment, not shown). Finally, a word was presented directly above the encoding question; participants had 3500 ms to respond. Retrieval trials began with presentation of one of three cues (“Side”, “Question”, or “Odd/Even”). After a 1000 ms delay, a word was presented. On Side and Question trials, the word came from the immediately preceding encoding block, while on Odd/Even trials the word was a numeral (e.g., “seventy-seven”). Finally, a response screen was presented and persisted until the participant responded or 10 seconds had elapsed. The response options for a Side trial are displayed. On Question trials, “left” and “right” were replaced with “living/non-living” and “mobile/immobile”, respectively; on Odd/Even trials they were replaced with “odd” and “even”.

*Figure 2*. Source memory (A) accuracy, (B, *top*) confidence, and (B*, bottom*) correct RT. Bar heights correspond to the mean, error bars = SEM. Asterisks denote *p* < 0.05.

*Figure 3*. Waveforms elicited by correct responses to the Question (black), Side (red), and Odd/Even (blue) cues. Representative electrodes from the left and right hemisphere are depicted for frontal and parietal scalp; a midline occipital electrode is displayed as this is where the late posterior negativity (LPN) was maximal. Gray shading demarcates the parietal ERP associated with recollection. Asterisks indicate the reduction in parietal activity in depressed adults.

*Figure 4*. Topographies showing activity elicited by Question and Side hits, with activity on correct Odd/Even trials subtracted out. Columns correspond to the three time windows analyzed (400-800, 800-1400, 1400-2000 ms). Electrodes in clusters associated with significant effects (within-group) are shown in white. Between-group comparisons revealed no differences.

*Figure 5*. Results of the mass univariate analysis conducted on Question minus Side difference waves, sorted by group and encoding task. Columns correspond to the three time windows analyzed (400-800, 800-1400, 1400-2000 ms). Electrodes in clusters associated with significant effects are shown in white. Paralleling the behavioral analyses, there were MDD > control differences in response to words from the mobility but not the animacy task.

*Figure 6.* ERP amplitudes are related to sleep in depressed adults. There were negative correlations between sleep disturbance as measured by the PSQI (*x*-axis) and ERP amplitudes captured by the Question minus Side difference wave for words from the mobility task (*y­*-axis). The correlations were significant in the 400-800 (left) and 800-1400 (right) ms time windows.