**An ERP Study of Source Memory in Unipolar Depression**

Elyssa M. Barrick and Daniel G. Dillon

Number of words:

Number of figures:

Number of tables:

Supplemental information: none

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:** The cognitive impairment associated with Major Depressive Disorder (MDD) may disrupt memory retrieval, but this hypothesis has rarely been tested. Therefore, we used event-related potentials to examine source memory in healthy and depressed adults.

**Methods:** Twenty-four adults with MDD and 24 controls encoded neutral words shown on the left or right (perceptual source) by making animacy or mobility judgments (cognitive source). The electroencephalogram was recorded as participants were cued to retrieve the perceptual and cognitive source of each word.

**Results:** Mobility judgments were associated with longer response times than animacy judgments, suggesting deeper encoding. Memory was characterized by a *Group* x *Cue* x *Encoding Task* interaction: depressed adults were less accurate and confident than controls, but they showed excellent cognitive source memory after mobility judgments. A positive ERP deflection over parietal cortex from 400-800 ms was reduced in depressed adults, but they showed sustained left parietal activation during cognitive source judgments for words from the mobility task.

**Conclusions:** Relative to controls, depressed adults showed modest reductions in source memory accuracy and confidence. Reduced left parietal activity from 400-800 ms suggests that impaired recollection is to blame. However, adults with MDD outperformed controls and showed sustained left parietal activation when cued to retrieve cognitive operations performed during deep encoding. This result echoes prior work showing excellent memory in depression provided attention is focused at encoding and retrieval, and the ERP data point to a candidate neural mechanism—sustained recruitment of parieto-hippocampal circuitry.

**Introduction**

A basic point from the source monitoring framework that’s worth keeping mind—source is not an item attribute, and we probably retrieve the memory first then reflect on it to inspect or deduce the source. The LPN may be tracking the latter process.

Another point: in their 1993 article on the source monitoring framework, Johnson, Hashtroudi, and Lindsay are very clear on the fact that source memories are not all-or-none and that we often remember certain aspects of a source but not others. I think what we have here is evidence that our encoding tasks made cognitive operations more vs. less accessible (probably because they were engaged to a greater extent; or more were engaged to a similar extent) at retrieval while having no differential effect on encoding of spatial position. Also, our LPN data show that recovering different aspects of a source elicits different (and sensible) patterns of brain activity (others have shown this, I think our data are especially clear). Ah, here’s a relevant quote from the SMF paper: “Because source monitoring depends on the information available from activated memory records, it relies fundamentally on the quality of the information recorded about events initially.” I wouldn’t push this point too hard but you can highlight the fact that many prior papers in this area have overlooked this point by looking at cue effects collapsed over encoding conditions (e.g., Bergrstom et al., 2013); we show that the Cue x Encoding Task interaction is robust.

Our task is not quite about internal-external monitoring because we specifically ask for either internal or external information. Instead, it’s about specifying which internal and external source information is from.

**Materials and Methods**

**Participants**

Participants were recruited from the community and compensated ($25.00/hour) for their time. All participants were 18-62 years old, right-handed, with no history of neurological or unstable medical conditions. Informed consent was obtained with a protocol approved by the Partners HealthCare Human Research Committee. To determine eligibility, prospective participants were screened by phone or online with an instrument that inquired about physical and mental health, medication use, and substance abuse. Controls had to report no current or past psychiatric conditions or unstable medical illness. Individuals were invited to participate in the MDD group if they reported current depression, no history of other DSM-IV Axis I diagnosis (excepting secondary generalized anxiety, social anxiety, or 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). Based on the screen, 34 healthy 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; Beck, Steer, & Brown, 1996). Data from depressed participants (*n* = 26) were retained if they met criteria for MDD but no other DSM-IV Axis I diagnosis with previously noted exceptions, and provided they had a BDI-II score ≥ 14. Data from healthy individuals (*n* = 34) were retained if they reported no current or past psychiatric illness. Finally, data from 2 depressed and10 healthy individuals were excluded due to excessive EEG artifacts (see section *EEG pre-processing*). Thus, the final sample consisted of 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 MASQ includes scales for anhedonic depression (MASQ-AD), anxious arousal (MASQ-AA), and general distress due to depression (MASQ-GDD) and anxiety (MASQ-GDA). The RRS provides measures of maladaptive (brooding) vs. adaptive (reflection) rumination, along with a scale that captures more general symptoms of depression. The PSQI assesses several sleep domains over the prior four weeks, including subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction. The PSQI also includes questions for the bed partner or roommate to complete, but these were not administered. Finally, the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) was used as a brief assessment of IQ. We included all these measures to characterize the MDD sample and to determine if any deficits associated with depression could be better understand as the consequence of a more narrowly defined process (e.g., brooding rumination, acute sleep disturbance).

**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.** Over the course of 100 encoding trials participants made animacy and mobility judgments for individual words. Therefore, we selected 100 words from the MRC Psycholinguistic Database (Coltheart, 1981), with 25 words from each of 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 per word (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. Emotionally neutral words were selected to avoid effects associated with mood congruent encoding or retrieval (G H Bower, 1981; Gordon H. Bower, 1987), thus isolating the cognitive impact of depression on retrieval. Word lists are printed in the appendix.

**Encoding.** The task included six encoding-retrieval cycles. Each encoding block consisted of 16 trials (Figure 1, *left*) in which a word appeared on the left or right side of the screen directly above one of two questions: “Living/Nonliving?” (animacy judgment) or “Mobile/Immobile?” (mobility judgment) (duration: 3.5 s). Participants responded by pressing a button corresponding to the correct answer. 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). Each block included four words from every category (living/immobile, non-living/immobile, living/mobile, non-living/mobile), each of which was assigned to one of the four encoding conditions defined by screen position and task (left/animacy, right/animacy, left/mobility, right/mobility). A jittered inter-trial interval (500-2000 ms) separated the encoding trials.

Immediately after each encoding block, a 3-digit number (262, 931, 888, 704, 557, or 474) was centrally presented for 30 s and participants counted backwards from that number, in steps of three, out loud. This task was used to disrupt sub-vocal rehearsal, clear working memory, and thus increase the difficulty of the upcoming retrieval test (Reitman, Higman, Lifson, & Rosenblum, 1974). To minimize stress, participants were told to strive for accuracy but, in case of a mistake, to simply proceed as though no error had been committed. The experimenter observed the counting but no data were collected during this task.

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 32 trials, the cue was “Side” or “Question” (16 trials each) and the word came from the immediately preceding encoding block. These cues prompted retrieval of the perceptual (“On what side did this word appear?”) and conceptual (“What question was answered for this word?’) 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: the participants had to read a cue, interpret it, and retrieve information from memory 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 and Question trials relative to Odd/Even trials should isolate activity due to episodic retrieval. Cues were printed directly above words, and 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. The numbers corresponded to response options indicating a choice and 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.

Each cue appeared for 1 s and was then joined by the word, which remained visible along with the cue for 3 s. At this point the response screen was presented and remained onscreen until the participant pressed a button or 10 s had elapsed. Cues were displayed before words because we anticipated time-locking the EEG data to the word onsets in order to study source retrieval. By presenting the cues first, we aimed to allow participants sufficient time to prepare a search strategy such that they could engage in retrieval as soon as each word appeared, increasing the likelihood that the ERP analysis would capture retrieval rather than preparation for retrieval. Similarly, we delayed the response screen in order to reduce the effect of motor preparation on the ERPs to words, and we allowed participants 10 seconds to respond in case of global slowing in MDD. Finally, a centrally presented fixation cross was continuously visible throughout retrieval, and the cues and words were presented directly above and below fixation, respectively, such that participants could see all the stimuli without needing to move their eyes. As at encoding, a jittered inter-trial interval (500-2000 ms) separated the retrieval trials. Very similar methods have been used in previous ERP and functional magnetic resonance (fMRI) imaging studies of cognitive versus perceptual source retrieval in healthy adults (Bergström, Henson, Taylor, & Simons, 2013; Simons, Owen, Fletcher, & Burgess, 2005; Simons, Gilbert, et al., 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 conventional ANOVAs because it is easy to include covariates that might have non-specific effects on source memory (e.g., age: Mark & Rugg, 1998) or depression (e.g., gender: Nolen-Hoeksema, 2001), thus improving the likelihood of detecting the unique effect of depression on source memory. The specific models used for each dependent variable are described below, but in all cases we computed an initial model with task elements (e.g., encoding task, source memory cue) and covariates (age, gender) as fixed effects but with *Group* omitted. Next, we added *Group* to the 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 test. 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, for which we modeled intercepts but did not adjust slope. When modeling encoding accuracy (coded 0 or 1), we used glmer with the logit link function. Finally, we extracted *p*-values for parameters using the R library *lmerTest*.

**Encoding**. Prior to analysis, we dropped trials with no response or where RT exceeded the participant’s mean±3SD; fewer than 1% of trials were dropped. We analyzed accuracy and RT. 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* as an additional fixed effect.

**Retrieval**. We dropped trials with no response or where RT exceeded the participant’s mean±3SD (fewer than 2% of trials). Next we analyzed the Odd/Even trials to judge their suitability as a control condition for the ERP analysis.As expected, 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* as factors, χ2s < 2.1, *p*s > 0.14, indicating that MDD did not affect performance. Thus, the Odd/Even trials elicited similar behavior across both groups and are therefore suitable as a control condition.

For the Side and Question trials, we conducted separate analyses for accuracy and confidence. 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 as follows: 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 from 0.1 to 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 from further analysis. 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 then time-locked to word onsets and segmented from 200 ms pre-stimulus to 2000 ms post-stimulus; the pre-stimulus interval was used for baseline correction. Segments where the maximum-minimum voltage difference (computed over 200 ms intervals with a 100 ms sliding window) exceeded 100 μV, or where any raw value exceeded +/- 100 μV, were rejected as artifacts.We rejected data from seven controls and two depressed adults where over 50% of trials were contaminated by artifact. Finally, the cleaned, segmented data were averaged for group-level analyses.

**Group-level analyses**. We first visually inspected the ERP waveforms and topographic maps associated with correct responses to the Question, Side, and Odd/Even cues, to verify the presence of the left parietal ERP associated with recollection, the right frontal ERP associated with monitoring, and the LPN. There was an obvious group difference in the left parietal ERP from 400-800 ms (see below), thus we extracted mean amplitude over this time window from four 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 is very similar to the method used in many prior studies (Hayama, Johnson, & Rugg, 2008; Rugg & Curran, 2007; Wilding & Rugg, 1996), and the focus on data from a small number of electrodes, over a single time window, constitutes the typical approach to ERP analysis.

Additional analyses used a more modern approach and were intended to parallel the behavioral 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 the groups (Groppe et al., 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 ERP analysis, mass univariate analysis involves a separate test at each electrode, either a one-sample *t*-test against zero (for within-group analysis) or a two-sample *t*-test. This is similar to the approach commonly used in fMRI research (Friston et al., 1995), and it makes much better use of the spatiotemporal richness of ERP data than the traditional method. However, appropriate correction for multiple comparisons can sharply limit power. To limit this concern, we used a cluster-based permutation procedure that can identify relatively weak effects so long as they extend over several electrodes (Groppe, Urbach, & Kutas, 2011b). Briefly,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, and we then performed 2500 permutations (within or across groups, depending on the test) to generate a null distribution of cluster mass values (Bullmore et al., 1999). Clusters whose mass probability was more extreme than all but 5% of the null distribution were considered reliable and are reported.

**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 the RT model, χ2s < 1.93, *p*s > 0.16, thus depressed and healthy adults performed similarly during encoding.

**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 significant 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 although accuracy in response to the Side cue did not vary substantially by encoding task, accuracy in response to the Question cue was better following mobility versus animacy judgments, and this effect appeared to be larger in depressed versus healthy adults. As shown in the right panel of Figure 2A, 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 difference 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 Question minus Side difference for the mobility task was significantly more positive in depressed versus healthy adults, *t*(46) = 3.04, *p* = 0.004, *d* = 0.88. Table 2 gives the percentage of responses in each cell of the design for completeness.

In summary, 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 a difference in favor of the MDD group for the Question/mobility cell of the design, *Z* = 1.98, *p* = 0.048, but in all other cells accuracy was non-significantly higher in the controls (*Z*s < 1.63, *p*s > 0.10). Finally, the best-fitting model also revealed effects of *Age*, *Z* = -3.31, *p* < 0.002, *Gender*, *Z* = 3.32, *p* < 0.002, and *Block, Z* = 3.23, *p* = 0.001. Thesereflected 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**. As shown in Figure 2B (top panel), 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, because the group difference in confidence was stronger under the Side cue, *Z* = 2.42, *p* = 0.016, than under 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 when responding to the Question vs. the Side cue, when responding to words encoded in the mobility vs. the animacy task, and when responding in later vs. earlier retrieval blocks.

**Source RT**. The bottom panel of Figure 2B shows that correct RT was similar across the groups, with all participants noticeably slower in response to the Question vs. Side cue. This impression was confirmed by the linear modeling, which was not improved by the addition of *Group*, *p =* 0.08, but which revealed 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.

**Summary**. Depressed adults were more accurate than controls when responding to words from the mobility task presented under the Question cue. At encoding, the mobility task was associated with longer reaction times and lower accuracy than the animacy task, suggesting that it elicited deeper processing (Craik & Tulving, 1975). Therefore, depressed adults performed best when cued to retrieve cognitive operations performed in the service of deep encoding. By contrast, they were (non-significantly) less accurate than the controls in every other cell of the design and they were less confident in all cells of the design, with this difference reaching significance under the Side cue. This amounts to a lower score for the MDD group in 7 of the 8 cells analyzed (8 cells: 2 *Cues* x 2 *Encoding Tasks* x 2 *Dependent Variables* [accuracy, confidence]) which is improbable under the null hypothesis of no group difference, binomial *p* = 0.035 (one-tailed). Thus, these data suggest a mild negative effect of MDD on source memory that can be corrected by combining especially deep encoding with a retrieval cue that directs attention to the cognitive operations performed at study (Hertel, 1997).

**ERPs**

Figures 3 and 4 display waveforms and topographic maps of correct responses to the Question, Side, and Odd/Even prompts, respectively. In controls, the three ERPs most frequently associated with episodic retrieval are readily apparent. From 400-800 ms there is robust activity over parietal electrode sites, and it is clear from the topographies that this effect is stronger over the left vs. right hemisphere; this potential has been consistently associated with recollection (Rugg & Curran, 2007). From 800-2000 ms, a positive potential is evident over right frontal cortex. This potential was initially thought to reflect post-retrieval monitoring but subsequent studies found that it can be elicited during semantic retrieval and decision-making more generally (Fleck, Daselaar, Dobbins, & Cabeza, 2006; Hayama et al., 2008; Hayama & Rugg, 2009). Consistent with this conceptualization, the late right frontal potential is evident even during Odd/Even judgments. Finally, a late negative potential (LPN) is evident during the same time window (800-2000 ms) as the right frontal effect (Cycowicz, Friedman, & Snodgrass, 2001). The LPN is not evident on Odd/Even trials, it has a medial posterior focus on Side trials, and it extends into left frontal cortex on Question trials.

The topographies from depressed adults are broadly similar but one difference was immediately apparent: parietal activity from 400-800 ms is markedly weaker in the MDD group. As outlined in the Methods section, we extracted the mean ERP amplitude between 400-800 ms from four parietal electrodes in the left and right hemispheres and submitted them to 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 activity in depressed adults. There was also a main effect of *Cue*, and follow-up tests using the REGWQ procedure yielded reliable differences between all three conditions (Question > Side > Odd/Even, all *ps* < 0.043).

PLEASE INSERT FIGURES 3 AND 4 ABOUT HERE

**Difference waves**. To isolate brain activity reflecting source retrieval, we created difference waves by subtracting activity on correct Odd/Even trials from activity elicited by hits in all four cells of the design (Question/animacy, Question/mobility, Side/animacy, Side/mobility), and then used the mass univariate approach to identify significant activations within and across the groups (Groppe et al., 2011a). In contrast to traditional ERP analysis, which typically involves conducting a few tests on a handful of electrodes specified *a priori*, mass univariate analysis involves a separate test at each electrode, conducted at each time point or on mean amplitude data over a pre-specified time window. This is very similar to the approach commonly taken in fMRI research (Friston et al., 1995), and it makes much better use of the spatiotemporal richness of ERP data than the traditional method. However, appropriate correction for multiple comparisons can sharply limit power. To limit this concern, we used a cluster-based permutation procedure that can identify relatively weak effects so long as they extend over several electrodes (Groppe et al., 2011b).

Briefly, for within-subject tests we submitted the mean amplitude of each difference wave over three time windows (400-800, 800-1400, 1400-2000 ms) to one-sample *t*-tests against zero. A separate *t*-test was computed at 123 electrodes (128 minus 5 electrodes placed on the face to monitor eye movements), andall electrodes within 4 cm of each other were considered neighbors. Neighboring electrodes with *p*-values less than 0.05 (uncorrected) were considered clusters. The sum of all *p­*-values within a cluster was considered its mass, and then 2500 within-subject permutations were performed to generate a null distribution of cluster mass values (Bullmore et al., 1999). Clusters whose mass probability was more extreme than all but 5% of the null distribution were considered reliable and are reported. Because the mass univariate toolbox does not currently support ANOVAs, after conducting within-group analyses we computed between-groups *t*-tests for each difference wave, again using the mass univariate approach with cluster-based permutation to correct for multiple comparisons.

**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, whichever came first. 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 highlights the effects of interest. Note the reduction in activity over parietal sites from 400-800 ms in depressed adults.

*Figure 4*. Topographies showing correct responses to the Question, Side, and Odd/Even cues. Note that parietal activity from 400-800 ms is blunted in depressed adults. In both groups, the LPN is confined to medial posterior sites on Side trials but extends over left frontal cortex on Question trials.