**Event-Related Potential Evidence of Disrupted Source Memory in Major Depressive Disorder**

Daniel G. Dillon and Elyssa M. Barrick

Motivated Learning and Memory Laboratory

Center for Depression, Anxiety and Stress Research

McLean Hospital

Harvard Medical School

Belmont, MA 02478

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:**

**Methods:** 24 controls and 24 unmedicated adults with MDD encoded words on the left or right side of a screen under one of two task prompts. At retrieval ERPs were recorded as participants were prompted to recall the location and task question associated with each word from encoding.

**Results:**

**Conclusions:**

**Introduction**

**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, and had no history of neurological disorders or unstable medical conditions. Informed consent was obtained with a protocol approved by the Partners HealthCare Human Research Committee. To determine eligibility for inclusion in the study, participants were screened over the phone or via a web-based instrument. The screen . . . [briefly describe the criteria used to select subjects; make sure to mention lack of medication for MDD] . . . Based on the screen, X healthy and Y 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 we administered the Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996). Data from depressed participants (*n* = X) were retained if they met criteria for MDD but no other DSM-IV Axis I diagnosis, with the exceptions of secondary generalized anxiety, social anxiety, or specific phobia, and provided they had a BDI-II score ≥ 14. Data from healthy individuals (*n* = Y) were retained if they reported no current or past psychiatric illness. Finally, data from X depressed and Y 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, participants completed the Mood and Anxiety Symptom Questionnaire(MASQ; Watson et al., 1995), the Ruminative Response Scale (RRS; Treynor, Gonzalez, & Nolen-hoeksema, 2003), and the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) (PSQI). The MASQ includes separate scales for . . . The RRS yields measures of brooding, reflection . . . The PSQI is a . . . Finally, the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) was used as a brief assessment of IQ. We included these measures to characterize the MDD sample and to determine whether any deficits associated with depression could be related to a more narrowly defined process (e.g., brooding rumination).

**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.** As described below, over the course of 100 encoding trials participants made animacy and mobility judgments for individual words. Therefore, we selected 100 words from **X** to serve as stimuli, 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) found no significant differences among the lists for the 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. Note that frequency, concreteness, and imageability data were not available for 12, 10, and 10 words, respectively. Importantly, we specifically selected emotionally neutral words as we wished to avoid any effects associated with mood congruent encoding or retrieval (G H Bower, 1981; Gordon H. Bower, 1987) in order to isolate the cognitive impact of depression on retrieval. The word lists are printed in the appendix.

**Encoding.** The task involved six encoding-retrieval cycles. Each encoding block consisted of 16 trials in which a word appeared on the left or right side of the screen directly above one of two questions: “Living/Nonliving?” or “Mobile/Immobile?” (duration: 3.5 s). Participants were told to respond to each question 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 or right) and a conceptual source defined by the encoding task (“Living/Nonliving?” vs. “Mobile/Immobile?”). 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 different encoding conditions defined by screen position and task (i.e., left/animacy, right/animacy, left/mobility, right/mobility). By counterbalancing the assignment of words according to category and encoding condition in this way, we sought to maximize control and minimize the possibility of spurious individual or group differences.

Immediately after each encoding block, a 3-digit number (262, 931, 888, 704, 557, or 474) was centrally presented and participants were asked to count backwards in steps of three, out loud, until the number was replaced by a fixation cross (30 s). The purpose of the counting was to disrupt sub-vocal rehearsal and thus increase the difficulty of the upcoming retrieval test (Reitman, Higman, Lifson, & Rosenblum, 1974). To minimize stress during counting, 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 participants to ensure that they engaged in backwards counting, but no data were collected during the counting task.

**Retrieval.** Each retrieval block comprised 48 trials that included a cue, a word, and a response screen. On 32 trials, the cue was either “Side?” or “Question?” (16 trials each) and the word was selected from the immediately preceding encoding block. Thus, these cues prompted the participant to retrieve perceptual (“On what side of the screen did this word appear?”) and conceptual (“What question did I answer for this word?’) information, respectively, for each encoded word. On the remaining 16 retrieval 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 were intended as a control condition: the participant must read the cue, interpret it, and retrieve information from memory before responding, exactly as in trials with the “Side?” and “Question?” cues. Critically, however, the “Odd/Even?” cue prompted retrieval from semantic rather than episodic memory. Therefore, comparing ERP data from the “Side?” and “Question?” conditions relative to the “Odd/Even?” condition should isolate activity specific to episodic retrieval. Cues were printed directly above the words.

The response screen consisted of the word ‘RESPOND’ printed above the word and the numbers 1-5 printed below. The numbers corresponded to response options that indicated the participant’s choice and his or her confidence in that choice: 1 = *high confidence* (left, living/non-living, odd), 2 = *low confidence* (left, living/non-living, odd), 3 = *guess*, 4 = *low confidence* (right, mobile/immobile, even), 5 = *high confidence* (right, mobile/immobile, even). Labels indicating the meaning of each response were printed below the numbers. Participants were instructed to select *guess* when they were unable to retrieve any information from encoding.

On every trial, the 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 give 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 our 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. A jittered inter-trial interval (500-2000 ms) separated the trials.

**Procedure.** Following application of the EEG net, participants were given detailed instructions and completed a practice encoding-retrieval cycle with four encoding trials and ten retrieval trials that included four “Side?” four “Question?” and two “Odd/Even?” cues. Thus, participants knew their memories would be tested prior to the first encoding block.

**EEG Recording**

The EEG was recorded 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 when possible; none exceeded 75 kΩ. EEG data were only acquired during retrieval.

**Behavioral Data Analysis**

In order to isolate effects of depression on retrieval it is necessary to account for many other factors that can influence memory, including age (Cabeza et al., 2004; Mark & Rugg, 1998), depth of encoding (Craik & Tulving, 1975), and the effects of different recognition cues (Konkel, Selmeczy, & Dobbins, 2015; Marsh & Hicks, 1998). Furthermore, because depression is more prevalent in women it is important to consider the possibility of gender effects (Nolen-Hoeksema, 2001). Therefore, 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), as linear mixed models can account for such potential confounds more easily than conventional ANOVAs while retaining sensitivity to the factors of interest. The specific models used for each dependent variable are described below, but in each case we first computed models with task elements (e.g., recognition cue) and covariates (age, gender) as fixed effects but with *Group* omitted. Next, we computed another model including *Group*, 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 on the first model, 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 from the best-fitting models using the R library *lmerTest*. Alpha was set to 0.05 throughout.

**Encoding**. Prior to statistical analysis, we dropped trials with no response or where the RT exceeded the participant’s mean±3SD; fewer than 1% of trials were dropped. We analyzed accuracy (did the participant answer the encoding question correctly?) and RT. Our first accuracy model included *Task* (“Living/non-living?” vs. “Mobile/immobile?”), *Side* (left, right), *Block* (1-6), *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 another fixed effect.

**Retrieval**. We removed trials with no response or in where the RT exceeded the participant’s mean±3SD; fewer than 2% of trials were removed. We also excluded the “Odd/Even?” trials as they were only included as a control condition for the ERP analysis. Before doing so, we examined behavior on these trials.Accuracy (percent correct) was at ceiling (controls: 98.43±0.12; MDD: 99.13±0.09), and RT (in ms) was similar between the 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, underscoring the fact that depression did not affect performance in this control condition.

For the remaining 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 treated as orthogonal to accuracy and was coded: guess = 1, low confidence = 2, high confidence = 3. We computed three models for both accuracy and confidence. The first included *Block*, *Cue* (“Side?” vs. “Question”), *Encoding Task* (“Living/Non-living?” vs. “Mobile/Immobile?”), *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 of *Group* and all two-way interactions involving *Group*. Finally, for retrieval RT we fitted the same models but included *Accuracy* and *Confidence* as additional factors. To simplify the RT analysis, we excluded guesses and coded *Accuracy* as “hit” or “miss” and *Confidence* as “high” or “low”. In summary, this approach to the retrieval data allowed us to determine how source memory was affected by the encoding task, the recognition cues, and their interactions with depression.

**EEG Analysis**

**Pre-processing.** Off-line analyses were conducted with the EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes for MATLAB (MathWorks, Natick). For each participant, the retrieval blocks were merged and re-referenced to the average of all electrodes prior to bandpass filtering from 0.1 to 30 Hz in EEGLAB. Bad channels were visually identified and interpolated. We set an *a priori* cut-off of up to 18 bad channels (14% of the total); datasets that exceeded this cut-off were excluded from further analysis (controls, *n* = **X**; MDD, *n* = **Y**). Next, data marked by gross artifacts were manually removed and independent component analysis was used to extract components capturing blinks, HEOG, and EKG. Using ERPLAB, the cleaned data were then time-locked to the word onsets and segmented from 200 ms pre-stimulus to 2000 ms post-stimulus; the pre-stimulus interval was also used for baseline correction. Next, a moving window peak-to-peak function flagged segments where the difference between the minimum and maximum voltage (computed over 200 ms intervals, 100 ms sliding window) exceeded 100 μV. Furthermore, any segment characterized by extreme values (+/- 100 μV) was also flagged. Finally, the data were visually inspected to ensure that all artifacts were appropriately marked, and then the artifacts were rejected. Any dataset with artifacts on more than 50% of trials was excluded from further analysis (controls, *n* = **X**, MDD, *n* = **Y**).The cleaned, segmented data were then averaged into bins defined by encoding position, encoding task, recognition cue, recognition accuracy, and confidence (e.g., encoded on left + animacy judgment + Side cue + recognition hit + high confidence).

**Group-level analyses**. Will pin down the behavioral results first as that will inform what gets presented here . . . \*DD note to self: complement the classic ERPs with trial-level analysis focused on 400-800 left parietal, maybe also right parietal (and right frontal + LPN)

**Results**

**Demographics**

As shown in Table 1, there were no group differences with regard to gender, age, or education. As expected, the MDD group endorsed more symptoms of 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 group difference in IQ as estimated from the WTAR.

PLEASE INSERT TABLE 1 ABOUT HERE

**Behavior**

**Encoding**. Encoding behavior was not affected by depression but it was influenced by the task: participants appeared to find the mobility judgment (“Mobile/Immobile?”) more difficult than the animacy judgment (“Living/Non-living?”). This was evident in both 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 also responded more quickly 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 (negative linear *Block* effect, *Z* = -6.34, *p* < 0.001), suggesting that fatigue was not an issue. Notably, the addition of *Group* did not improve the accuracy or the RT model, χ2s < 1.93, *p*s > 0.16. Thus, depressed and healthy adults performed similarly, and the mobility judgment was more difficult than the animacy judgment.

**Recognition accuracy**. As shown in Figure 1A, depressed adults were less accurate than controls except in response to words encoded in the mobility task and presented under the “Question?” cue. The figure also depicts an interaction between retrieval cue and encoding task: under the “Question?” cue, participants responded more accurately to words from the mobility vs. animacy task, whereas responses to the “Side?” cue were not sensitive to the encoding task.

These impressions were confirmed by the linear models, which were improved by the addition of *Group*, χ2 = 26.40, *p* < 0.001. The best-fitting model included a *Group* x *Cue* x *Encoding Task* interaction, *Z* = -2.13, *p* = 0.033, which subsumed significant *Cue* x *Task* and *Group* x *Task* interactions as well as significant main effects of *Cue* and *Task* (*Z*s > 2.7, *p*s < 0.006). Two sets of follow-up analyses unpacked the interaction. First, pairwise comparisons confirmed that depressed adults were more accurate than controls when responding to words encoded in the mobility task and presented under the “Question?” cue, *Z* = 1.98, *p* = 0.048. In all other cells the controls outperformed the MDD group, although the difference was never significant (*Z*s < 1.63, *p*s > 0.10). This pattern is consistent with a modest negative effect of depression on source memory that faded when a conceptually-oriented cue (“Question?”) was paired with stimuli from a deep encoding condition (mobility judgment).

Second, breaking down the data by cue type revealed a significant *Group* x *Task* interaction for responses to the “Question?” cue, *Z* = 3.25, *p* = 0.001: both groups responded more accurately to words from the mobility vs. animacy task, but the difference was larger in depressed (*Z* = 9.21, *p* < 0.001) vs. healthy (*Z* = 4.62, *p* < 0.001) participants, leading to the interaction. By contrast, the *Group* x *Task* interaction was not significant for responses to the “Side?” cue, *Z* = 0.57, *p* = 0.57. Thus, participants were able to retrieve perceptual information equally well regardless of the encoding task, but their ability to retrieve conceptual information (i.e., which question was answered at encoding) was better for words encoded in the context of the mobility task vs. the animacy task, with this last effect more pronounced in the MDD group.

Finally, there were significant effects of *Age*, *Z* = -3.31, *p* < 0.002, and *Gender*, *Z* = 3.32, *p* < 0.002, as well as a linear *Block* effect*, Z* = 3.23, *p* = 0.001. Thesereflected higher accuracy in younger vs. older adults, in men vs. women, and in later vs. earlier blocks, respectively.

PLEASE INSERT FIGURE 1 ABOUT HERE

**Recognition confidence**. The confidence data shown in Figure 1B indicate the depressed adults were less confident than controls, with the difference most pronounced under the “Side?” cue. 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. The interaction reflected the fact that while the controls were generally more confident than the depressed adults, the group difference 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 strong effects of *Cue*, *Z* = -5.33, *p* < 0.001 and *Encoding Task*, *Z* = 2.91, *p* = 0.004, as well as a positive linear effect of *Block*, *Z* = 3.69, *p* < 0.001. These results reflect the fact that participants were more confident when responding to the “Question?” cue vs. the “Side?” cue, when responding to words encoded in the mobility task vs. the animacy task, and when responding in later vs. earlier retrieval blocks. In contrast to accuracy, confidence was not affected by *Age* or *Gender*.

**Recognition RT**. Figure 1C shows that RT was similar across the groups with one exception: relative to controls, depressed adults responded more slowly to words from the animacy task presented under the “Question?” cue. Furthermore, all participants were noticeably slower in response to the “Question?” vs. “Side?” cue. These impressions were partially confirmed by the linear modeling, which was improved by the addition of *Group*, χ2 = 11.62, *p* = 0.020, and included a significant *Group* x *Cue* x *Encoding Task* interaction, *Z* = 2.42, *p* = 0.016. This triple interaction subsumed *Group* x *Cue* and *Group* x *Task* interactions, as well as a main effect of *Cue* (*Z*s > 2.50, *p*s < 0.013). Despite the interactions, and in contrast to the impression from the figure, pairwise comparisons did not show significantly slower responses for depressed vs. healthy participants in response to words from the animacy task presented under the “Question?” cue, or for any of the other cue/encoding task combinations (*Z*s < 1.05, *p*s > 0.29). However, the effect of *Cue* was exceptionally robust, *Z* = -23.91, *p* < 0.001, reflecting consistently longer RTs in response to the “Question?” vs. “Side?” cue across both groups.

Finally, the best-fitting model also returned significant main effects of *Accuracy* and *Confidence*, as well as a negative linear tread for *Block* (*Z*s > 3.08, *p*s < 0.004). Participants responded more quickly when accurate vs. inaccurate, when they were more vs. less confident, and during later vs. earlier blocks. There was also a main effect of *Gender*, as males responded more quickly than females.

**Summary**.