**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 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 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, we administered 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 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.** As described below, over the course of 100 trials participants made animacy and mobility judgments for individual words during encoding. Therefore, we selected 100 words (**give the source of the words**) to serve as stimuli, with 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 affective 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 included 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 (i.e., “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 their category and encoding condition, we attempted to maximize control and minimize the likelihood 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 difficult 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 this part of the 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”, inclusive, and the participant was asked to judge the parity of the numeral. “Odd/Even?” trials were intended as a control condition: note that on these trials the participant must read the cue, interpret it, and retrieve information from memory before responding, exactly as in trials beginning 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 until 10 s elapsed. The rationale for displaying cues before words was that we anticipated time-locking the EEG data to the word onset times 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, thus reducing contamination of the EEG with EOG signals. 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**

Our goal was to investigate whether depression affects retrieval, but it is clear that retrieval is sensitive to many other factors including age (Cabeza et al., 2004; Mark & Rugg, 1998), depth of encoding (Craik & Tulving, 1975), and the presence of different recognition cues (Konkel, Selmeczy, & Dobbins, 2015; Marsh & Hicks, 1998). To isolate effects of depression while accounting for these factors, we analyzed the retrieval (and encoding) data using linear mixed models implemented with the R (R Developement Core Team, 2015) library *lme4* (Bates, Maechler, Bolker, & Walker, 2015). Specific models used for the different dependent variables are described below, but in all cases we first computed models with task elements (e.g., recognition cue) and covariates of no interest (age, gender) as fixed effects but with *Group* omitted. Next, we computed another model after adding *Group*, either solely as a main effect or in interaction with other factors, and then we used likelihood ratio tests implemented in the R library *anova* to compare model fits by chi-square test. 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*.

**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. 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 dropped trials with no response or in where the RT exceeded the participant’s mean±3SD; fewer than 2% of trials were dropped. 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 was at ceiling (percent correct: controls = 98.43±0.12; MDD = 99.13±0.09), and RT was similar between the groups (RT (ms): 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 “Side?” and “Question?” trials, the accuracy of each response 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* as factors. 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*; otherwise this model was identical to the second model. 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”.

**EEG Analysis**

**Pre-processing.** Off-line analyses were performed using EEGlab and ERPlab toolboxes for Matlab. The data blocks were merged and re-referenced to the average of all the scalp electrodes prior to bandpass filtering of 0.1 Hz and 30 Hz. Channels with excessive artifacts were manually identified and interpolated, with a maximum number of 18 channels (14%). Data were then visually inspected for gross artifacts and sections manually rejected. Independent component analysis was used to identify components containing eye blinks, eye movement, and ECG artifacts, which were manually chosen for removal. Trials were analyzed for extreme artifact values using simple voltage threshold set at -100 to 100 μV. Participants were excluded if more than 50% of trials were rejected. For each participant, the data were epoched into 200 ms pre-stimulus, 2000 ms post-stimulus trials that were time locked to the onset of the prompt cue (side, question, odd/even). ERPs were generated at the subject level and grand averaged within group.

\*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 respect to gender, age, or education. As expected, the MDD group endorsed more symptoms of depression and anxiety than the controls did, with the mean BDI-II score in the MDD group indicating moderate depression. The MDD group also reported more rumination and poorer sleep than controls, but there was no group difference in IQ as estimated from scores on the WTAR.

**Behavior**

**Encoding**. Encoding behavior was not affected by depression but it was influenced by the task: participants found the “Living/Non-living?” question (animacy judgment) easier than the “Mobile/Immobile?” question (mobility judgment). This was evident with respect to accuracy (percent correct: animacy = 95.85±0.20; mobility = 92.42±0.26; *Z* = 4.91, *p* < 0.001) and RT (animacy = 1,664±535 ms; mobility = 1,801±552 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 course of the experiment (linear trend for *Run*, *Z* = -6.34, *p* < 0.001). The addition of *Group* did not significantly improve either the accuracy or the RT model, χ2s < 1.93, *p*s > 0.16.

**Recognition accuracy**. This should be Figure 1 . . .

**Recognition confidence**.

**Recognition RT**.