**Depression Modulates Brain Activity During Multidimensional Source Retrieval**

Elyssa M. Barrick and Daniel G. Dillon

Center for Depression, Anxiety and Stress Research, McLean Hospital

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**

Major Depressive Disorder (MDD) is associated with impaired episodic retrieval, but few imaging studies have examined this issue. Therefore, we collected event-related potentials (ERPs) from 24 unmedicated adults with MDD and 24 healthy controls during source memory decisions. At encoding, neutral words were presented on the left or right (perceptual source) in the context of animacy or mobility judgments (conceptual source). Mobility judgments were made more slowly than animacy judgments, consistent with sustained sematic analysis. At retrieval, accuracy for words from the mobility task was characterized by a *Group* x *Cue* interaction, with the MDD group showing a benefit for conceptual vs. perceptual retrieval that was not observed in controls. Similarly, left parietal ERPs elicited by words from the mobility task were characterized by *Group* x *Cue* interactions from 400-800 ms and 800-1400 ms post-stimulus. In these intervals, the MDD group showed higher amplitude ERPs for conceptual vs. perceptual retrieval, but the controls did not. In the MDD group, the magnitude of these ERP effects was negatively correlated with self-reported sleep disruption. By contrast, conceptual source accuracy was worse than perceptual source accuracy for words from the animacy task across both groups, and an ERP difference score tracking this effect was characterized by sustained negative potentials over fronto-central scalp. These data highlight the sensitivity of conceptual (but not perceptual) source memory to the extent of semantic processing at encoding, and indicate that source accuracy in MDD covaries with left parietal ERPs linked to recollection.

Keywords: source memory, retrieval, ERP, recollection, depression

**Introduction**

Major Depressive Disorder (MDD) is known to bias emotional memory. Relative to healthy controls, adults with MDD typically show excellent memory for negative stimuli and poor memory for positive stimuli (Burt, Zembar, & Niederehe, 1995; Daniel G Dillon, Dobbins, & Pizzagalli, 2014; Hamilton & Gotlib, 2008; Matt, Vazquez, & Campbell, 1992), and we have argued that the positive memory deficit may reflect disruption of brain dopamine systems (Daniel G Dillon, 2015). Importantly, though, one need not use emotional stimuli to detect an episodic memory deficit in MDD. For instance, Zakzanis and colleagues (1998) performed a meta-analysis of 22 neuropsychological studies and found strong negative effects of depression on episodic memory tests that used neutral stimuli. Similarly, Airaksinen and colleagues (2004) used neutral words to demonstrate impaired free and cued recall in a sample of 187 depressed adults from the community. Finally, Rock and colleagues (2014) analyzed over 30 years of work with the Cambridge Neuropsychological Test Automated Battery and found that, relative to healthy adults, depressed individuals showed significantly impaired episodic memory, despite the fact that all the tasks used neutral stimuli.

What underlies these memory impairments? One possibility is poor encoding. For instance, Airaksinen et al. (2004) found that although memory accuracy was lower in depressed vs. healthy adults, both groups showed a similar benefit for cued vs. free recall, implying that the depressed group was able to use the cues to enhance retrieval but simply had stored less material in the first place. A more nuanced result was presented by Zakzanis et al. (1998), who noted that the effect of depression on memory was larger for tasks that provided less support during learning (e.g., uncategorized vs. categorized word lists). Hertel’s *cognitive initiative* framework offers a principled account of such findings (Paula T. Hertel, 1997; Paula T. Hertel & Hardin, 1990). The framework’s core hypothesis is that depressed individuals can muster the resources to control attention and use effective strategies at encoding but, in the absence of external support or emotionally compelling material, they often fail to do so. To test this hypothesis, Hertel and Rude (1991) presented neutral words and sentence frames to depressed and healthy participants, who were asked to judge whether each word fit its frame. Critically, participants were assigned to either focused or unfocused conditions. In the focused condition, each word was flashed briefly and was then replaced by its frame for 8 s, at which point the participant restated the word and indicated whether or not there was a quality fit. In the unfocused condition, the words and frames were presented concurrently, the participant was not required to restate each word, and responses could be made at any time. Because the unfocused condition was designed to permit mind-wandering and rumination, both of which are common in depression (e.g., Nolen-Hoeksema, 1991), and because mind-wandering and rumination were expected to detract from encoding, Hertel and Rude predicted a negative effect of depression only in the unfocused condition. Indeed, free recall was characterized by a *Group* x *Task* interaction, with depressed adults remembering fewer words than controls after the unfocused task but more words than controls after the focused task. Thus, tasks that engage attention and encourage elaborative processing at encoding can reduce the negative effect of depression on memory.

Importantly, the cognitive initiative framework applies to retrieval as well. Depressed adults typically show larger deficits for free recall than for recognition (Burt et al., 1995), and when recognition memory is analyzed to estimate the contributions made by recollection vs. familiarity (or controlled vs. automatic processing), depression impairs the former more than the latter (e.g., Paula T. Hertel & Milan, 1994; G. M. MacQueen, Galway, Hay, Young, & Joffe, 2002). The cognitive initiative framework explains these data by pointing to the greater need for controlled attention and generation of search strategies during free recall vs. recognition, and in support of recollection vs. familiarity. To test this account, Hertel and Milan (1994) conducted studies in which healthy and depressed participants—either naturally depressed or temporarily “depressed” based on a negative mood induction—completed successive implicit (homophone spelling) and explicit (recognition) memory tests for the same words. When the experimenters did not draw attention to the fact that the same words appeared on both tests, healthy participants were more likely than depressed adults to notice this regularity and exploit it to improve recognition memory (i.e., by referring back to the implicit test). However, once the dependency between the tests was highlighted, the depression-related deficit in recognition memory disappeared. Moreover, when participants were given a detailed strategy for improving memory, depressed adults showed an especially strong benefit. These results again support the idea that depression decreases the spontaneous use of strategies that can improve memory, but that the capacity to benefit from such strategies is unaffected.

Given the extensive and elegant behavioral work on these issues, the paucity of data regarding the relevant neurobiology is surprising. In particular, although there are many studies focused on hippocampal volume reductions in MDD (for review, see G. MacQueen & Frodl, 2011) and some functional imaging studies of encoding (e.g., Bremner, Vythilingam, Vermetten, Vaccarino, & Charney, 2004; D.G. Dillon & Pizzagalli, 2013; Daniel G Dillon et al., 2014; Hamilton & Gotlib, 2008), there are remarkably few neuroscientific studies of episodic retrieval in MDD. This does not reflect lack of desire—a decade ago, the National Institutes of Mental Health, Aging, and Neurological Disorders and Stroke called for integrated neuroscientific research on depression and memory (Steffens et al., 2006). However, despite dozens of event-related potential (ERP) and functional magnetic resonance imaging (fMRI) studies of episodic retrieval in healthy adults (Eichenbaum, Yonelinas, & Ranganath, 2007; Rugg & Curran, 2007; Rugg & Vilberg, 2013), no similar literature in MDD has emerged.

The current ERP study was designed to address this gap. In light of the fact that depression more strongly affects recollection than familiarity—and given the difficulties associated with imaging free recall—we elected to study source memory in MDD. Source memory refers to conscious retrieval of the spatiotemporal details that define an encoding episode (Johnson, Hashtroudi, & Lindsay, 1993); it depends heavily on recollection, and there is evidence that it is disrupted in depression (Degl’Innocenti & Bäckman, 1999). We adapted a design that dissociates neural systems engaged by conceptual vs. perceptual source retrieval (Bergström, Henson, Taylor, & Simons, 2013; Dobbins & Wagner, 2005; Simons, Gilbert, et al., 2005). At study, participants viewed neutral words presented on the left or right above a question specifying either an animacy or mobility judgment. At test, they were cued to retrieve the presentation side (perceptual source) and encoding task (conceptual source). A recent fMRI/ERP study (Bergström et al., 2013) found that both conceptual and perceptual retrieval elicited the most well-studied ERP marker of recollection: a positive deflection over parietal cortex that extends from about 400-800 ms post-stimulus, typically with a left hemisphere maximum (Rugg & Curran, 2007). Both forms of retrieval also activated the precuneus and elicited a negative polarity ERP maximal over posterior electrodes that is referred to as the late posterior negativity, or LPN (Cycowicz, Friedman, & Snodgrass, 2001; Johansson & Mecklinger, 2003; Mecklinger, Johansson, Parra, & Hanslmayr, 2007). Importantly, the LPN extended over left frontal cortex during conceptual retrieval, and this was mirrored by fMRI activation in the dorsolateral PFC. Related fMRI studies confirmed that left PFC and medial PFC were more strongly activated during conceptual vs. perceptual source retrieval (Simons, Gilbert, et al., 2005; Simons, Owen, Fletcher, & Burgess, 2005).

At the outset of the study, we hypothesized that depression would selectively impair conceptual source memory. This *a priori* hypothesis was based on two ideas. First, MDD is associated with volumetric loses in medial PFC (Treadway et al., 2015), and to the extent that medial PFC is especially important for conceptual vs. perceptual retrieval (Simons, Gilbert, et al., 2005; Simons, Owen, et al., 2005), we anticipated that conceptual retrieval would be especially disrupted by depression. Second, rumination is negatively associated with episodic memory in depression (Paula T. Hertel & T., 1998), and rumination activates the PFC circuits that are particularly active during conceptual source retrieval (Cooney, Joormann, Eugène, Dennis, & Gotlib, 2010). Thus, we expected depressive rumination to impair conceptual source retrieval. Because prior work using this experimental paradigm had not emphasized interactions between the encoding tasks and either conceptual or perceptual retrieval—indeed, the retrieval data have typically been presented collapsed over different encoding tasks in prior studies—we did not expect variation in retrieval performance based on the encoding tasks in either group.

However, it quickly became apparent that this expectation was in error. As described below, although both encoding tasks required analysis of the semantic properties of the words, the mobility task was more difficult and elicited longer response times (RTs). This had consequences for conceptual but not perceptual retrieval across both groups, and—consistent with the cognitive initiative framework—we found that the combination of sustained, focused analysis at encoding with conceptual retrieval led to surprisingly good performance in the MDD group. Thus, this experiment provided an unforeseen opportunity: we were able to take a first look at neural activity that mediates accurate source memory in depression when both encoding and retrieval are well-supported.

**Materials and Methods**

**Participants and self-report**

Participants (18-62 years old, right-handed, no neurological or unstable medical conditions) were recruited from the community and compensated $25/hour, following a protocol approved by the Partners HealthCare Human Research Committee. Participants were screened by phone or online, at which time the Beck Depression Inventory II (BDI; Beck, Steer, & Brown, 1996) was administered. Prospective participants were invited to participate in the MDD group provided they endorsed symptoms consistent with a current Major Depressive Episode, had a BDI-II score ≥ 14 (the accepted cut-off for mild depression; Beck et al., 1996), and reported no other Axis I psychopathology with the exception of generalized anxiety, social anxiety, and/or specific phobia. Participants in the control group had to report no current or past Axis I psychopathology. On the day of the ERP experiment, we conducted a detailed assessment of psychiatric status by administering the MINI International Neuropsychiatric Interview, version 6.0 (Sheehan et al., 1998). Depressed adults had to again report current depression, no history of other DSM-IV Axis I diagnosis (except generalized anxiety, social anxiety, or specific phobia secondary to MDD), and no medication use in the past two weeks (six weeks for fluoxetine, six months for neuroleptics). Thirty-four controls and 26 depressed adults completed the ERP session. Data from 10 controls and 2 depressed adults were excluded due to excessive artifacts (see below), leaving 24 individuals per group.

Following the EEG session, we administered the BDI-II again, along with 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). These probe symptoms of depression and anxiety, trait rumination, and sleep quality over the last month, respectively. The BDI, MASQ, and RRS are commonly used to assess cognitive and affective aspects of depression, with the MASQ also providing insight into anxiety. The PSQI was included because sleep has beneficial effects on human episodic memory (Plihal & Born, 1997) and there is substantial evidence linking depression and other psychiatric disorders to sleep disruption (Deldin, Phillips, & Thomas, 2006; Wulff, Gatti, Wettstein, & Foster, 2010). Thus, we expected to find negative relationships between these measures on one hand and both memory accuracy and ERP indices of successful source retrieval on the other. Finally, the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) was used to estimate IQ. One control did not complete the MASQ and 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 recorded for one control and one depressed participant.

**Stimuli.** 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*). ANOVA yielded no differences for number of letters (mean±S.D.; 5.27±1.29) or syllables (1.52±0.50), frequency (35.58±79.02), concreteness (598.87±20.18), or imageability (596.80±25.31), *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 above one of two questions: “living/non-living?” or “mobile/immobile?” Participants responded by pressing a button. A jittered interval (500-2000 ms) separated the trials.

**Counting.** Immediately after encoding, a 3-digit number (e.g., 931) was shown and participants counted backwards from that number in steps of three for 30 s. Counting served to disrupt rehearsal and clear working memory (Reitman, Higman, Lifson, & Rosenblum, 1974).

[PLEASE INSERT FIGURE 1 ABOUT HERE]

**Retrieval.** Each 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 preceding encoding block; these cues prompted perceptual and conceptual source retrieval, respectively. On the remaining trials the cue was “Odd/Even” the word was a numeral between “one” and “ninety-six”, and the participant judged parity. All trials involved reading a cue, interpreting it, and retrieving information, but on Odd/Even trials retrieval was directed at semantic rather than episodic memory. Thus, comparing ERP data from Side or Question trials vs. Odd/Even trials should isolate activity mediating episodic retrieval. Presentation order of words and cues was random. The response screen consisted of ‘RESPOND’ printed above the word with the numbers 1-5 printed below and corresponding to a choice and level of confidence (Figure 1, *right*). As in other studies of source memory (Starns & Hicks, 2005a), “guess” was included as a response option. Participants were asked to select “guess” when they could not recover any information favoring one source over the other, so that an analysis focused on hits should not be contaminated with guesses. A jittered interval (500-2000 ms) separated the trials.

**EEG Recording**

The EEG was recorded during retrieval with a 128-sensor HydroCel GSN Electrical Geodesics Inc (EGI) net (sample rate: 1000 Hz, 0.02–100 Hz). Data were referenced to vertex and impedances were kept below 45 kΩ when possible (maximum: 75 kΩ).

**Behavioral Analysis**

The behavioral data were cleaned by dropping trials with no response or where RT exceeded the participant’s mean±3SD (< 1% of encoding trials, < 2% of retrieval trials). Analysis involved *t*-tests andmixed-model Type III ANOVAs implemented in the R software (R Developement Core Team, 2015) library *afex* (Singmann, Bolker, Westfall, & Aust, 2016). For both encoding and retrieval, accuracy was computed as percent correct. At encoding, *Group* x *Task* (mobility, animacy) x *Side* ANOVAs were run for accuracy and correct RT. At retrieval, *t­*-tests were first used to compare accuracy, confidence, and correct RT on Odd/Even trials, to verify that there were no group differences in this control condition. Next, a *Group* x *Cue* x *Task* (mobility, animacy) ANOVA was run on the number of guesses in each condition. Finally, responses on Question and Side trials were analyzed by running *Group* x *Cue* x *Task* ANOVAs for accuracy, confidence, and correct RT. Alpha was set at 0.05.

**ERP Analysis**

**Pre-processing.** Pre-processing was conducted with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes for MATLAB (MathWorks, Natick). EEG data were merged, re-referenced to the average of all electrodes, and filtered (0.1-30 Hz). Bad channels were interpolated, independent components analysis was used to remove activity reflecting blinks, HEOG, and EKG, and the cleaned data were time-locked to word onsets and segmented (-200 to 2000 ms). The pre-stimulus interval was used for baseline correction, and segments where any raw value or the maximum-minimum voltage difference (200 ms intervals, 100 ms sliding window) exceeded 100 μV were rejected. We used *a priori* criteria of > 18 bad channels or more than 50% of trials rejected (Luck, 2014) to exclude datasets (10 controls, 2 MDD). 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 and there were too few clean segments for analyzing misses. Thus, the analysis was focused on correct responses, a common approach in this literature (Bergström et al., 2013; Dobbins & Wagner, 2005; Han, Oʼconnor, Eslick, & Dobbins, n.d.; Simons, Gilbert, et al., 2005).

**Group-level analyses**. We conducted two group-level analyses. In the first, we computed “Question minus Odd/Even” and “Side minus Odd/Even” difference waves to isolate activity tracking conceptual and perceptual source retrieval, respectively (Bergström et al., 2013). These contrasts were collapsed over the encoding tasks and permitted a test of our *a priori* prediction that depressed adults would selectively affect conceptual retrieval. The second analysis was intended to parallel the source accuracy results, which revealed a *Group* x *Cue* interaction for words from the mobility task but only a main effect of *Cue* for the animacy task. To isolate the cue effect for each task, we computed “Question minus Side” difference waves separately for words from the mobility and animacy tasks in each group, and then we compared the groups to look for *Group* x *Cue* interactions for each task. For both group-level analyses, we submitted the difference waves to mass univariate analysis (Groppe, Urbach, & Kutas, 2011a), focusing on mean amplitudes from 400-800 ms, 800-1400 ms, and 1400-2000 ms. The 400-800 ms interval was selected to capture the left parietal effect consistently associated with recollection (Rugg & Curran, 2007), with the latter two windows selected to capture the LPN consistently seen during source recollection (Bergström et al., 2013; Johansson & Mecklinger, 2003).

Mass univariate analysis is widely used in fMRI research (Friston et al., 1995) and here entails a one-sample (within-group analysis) or a two-sample (between-group analysis) *t*-test at each electrode. By examining every electrode and multiple time windows, this approach makes better use of the spatiotemporal richness of ERP data than traditional methods. To correct for multiple comparisons, we used cluster-based permutation (Groppe, Urbach, & Kutas, 2011b). All electrodes within 4 cm of each other were considered neighbors, and neighboring electrodes significant at *p* < 0.05 (uncorrected) were considered clusters. The use of a 4 cm inter-electrode distance appropriate for detecting neighboring electrodes in regions where electrodes are relatively widely spaced (e.g., over parietal cortex) on the 128 channel EGI net (Song et al., 2015). The sum of all *p*-values in a cluster constituted its mass. We then performed 2500 between-participant permutations of the data, each time randomly assigning participants to the two groups without replacement and then selecting the most extreme cluster mass from each permutation to generate a distribution for judging the probability of observing clusters of various sizes under the null hypothesis of no difference between the groups (Bullmore et al., 1999). This null distribution was then used to judge the probability of observing the clusters actually observed when controls and depressed adults were compared. Only clusters significant at *p* < 0.05 (corrected) are reported.

**Individual Differences**

Across the groups, we used Pearson correlations to examine relationships between source memory accuracy, confidence, and left parietal ERPs associated with recollection. Within the depressed group, we examined relationships between source memory accuracy, left parietal ERPs, sleep quality, the severity of depressive and anxious symptoms, and brooding rumination.

**Results**

**Demographics**

There were no group differences in gender, age, education, or estimated IQ (Table 1). Relative to controls, the MDD group endorsed poorer sleep, more rumination, and more symptoms of depression and anxiety, with the mean BDI-II score indicating moderate depression.

[PLEASE INSERT TABLE 1 ABOUT HERE]

**Behavior**

**Encoding**. The encoding data are presented in Figure 2. For response accuracy (Figure 2A), the only significant result was a main effect of *Task*, *F*(1, 46) = 12.70, *p* < 0.001, *d* = 0.46, reflecting lower accuracy for mobility vs. animacy judgments. Neither the main effect of *Group* nor any interaction involving this factor approached significance, *F*s < 1. For correct RT (Figure 2B), the main effect of *Task* was again significant, *F*(1, 44) = 54.34, *p* < 0.001, *d* = 0.30, with slower responses for mobility vs. animacy judgments. The RT analysis also revealed a *Group* x *Task* x *Side* interaction, *F*(1, 44) = 6.40, *p* = 0.015, but separate ANOVAs for words presented on the left and right did not reveal significant *Group* x *Task* interactions, *F*s < 3.69, *p*s > 0.05. Examining Figure 2B, the 3-way interaction appears to reflect the fact that while RTs were consistently numerically shorter in the MDD group, this difference was more pronounced for the animacy task for words shown on the left, and for the mobility task for words shown on the right. In summary, the mobility task was more difficult than the animacy task, as judged by lower response accuracy and slower RTs, and there were no reliable group differences.

[PLEASE INSERT FIGURE 2 ABOUT HERE]

**Retrieval: Odd/Even trials**. There were no group differences on Odd/Even trials, which were characterized by extremely high accuracy (controls: 98.42±3.96%; MDD: 99.13±1.36%; *t*(46) = -0.83, *p* = 0.41), fast correct RTs (controls: 859.87±299.18 ms; MDD: 777.22±225.65 ms; *t*(44) = 1.06, *p* = 0.30), and highly confident responses (percentage of high confidence Odd/Even trials: controls: 99.69±0.65%; MDD: 99.87±0.35%; *t*(46) = -1.16, *p* = 0.25).

**Retrieval: guessing**. Five participants (three controls, two MDD) never guessed and a further 15 participants (ten controls, five MDD) did not guess in at least one cell of the design. Data from the remaining participants (11 controls, 17 MDD) are shown in Figure 3 and were included in the *Group* x *Cue* x *Task* ANOVA; because no participant guessed on an Odd/Even trial, that condition was omitted. The ANOVA yielded an effect of *Cue*, *F*(1, 26) = 7.11, *p* = 0.01, reflecting fewer guesses under the Question vs. Side cue, and an effect of *Task*, *F*(1, 26) = 12.34, *p* = 0.002, reflecting fewer guesses to words from the mobility task vs. the animacy task.

[PLEASE INSERT FIGURE 3 ABOUT HERE]

Neither the main effect of *Group* nor any interactions involving *Group* were significant, *F*s < 1.83, *p*s > 0.18. However, inspection of Figure 3 suggested that, for words from the mobility task, the effect of cue type was pronounced in the MDD group. Indeed, a *t*-test confirmed that for the mobility task, depressed participants guessed less frequently under the Question vs. the Side cue, *t*(16) = -3.11, *p* = 0.007, *d* = 0.51; the cue effect was not reliable for words from the animacy task considered alone, *t*(16) = -1.29, *p* = 0.22. In controls there was no reliable cue effect on guessing in either encoding task considered alone (*t*s < 1.33, *p*s > 0.21). Moreover, in the MDD group, words from the mobility task presented under the Question cue elicited fewer guesses than words from the animacy task shown under the Question cue, *t*(16) = -3.00, *p* = 0.008, *d* = 0.38, or words from the animacy task shown under the Side cue, *t*(16) = -3.09, *p* = 0.007, *d* = 0.75. In summary, participants guessed less to words from the mobility task vs. the animacy task, and in response to the Question cue vs. the Side cue. In the MDD group, the combination of these factors—words from the mobility task presented under the Question cue—led to particularly few guesses.

**Retrieval: source accuracy**. The source accuracy data are shown in Figure 4A. These were characterized by main effects of *Cue*, *F*(1, 46) = 17.47, *p* < 0.001, and *Task*, *F*(1, 46) = 31.95, *p* < 0.001, that were qualified by two interactions: *Cue* x *Task*, *F*(1, 46) = 22.35, *p* < 0.001, and *Group* x *Cue*, *F*(1, 46) = 6.42, *p* = 0.01. The *Cue* x *Task* interaction reflected the fact that when the data were collapsed across the two groups, accuracy was significantly lower under the Question vs. Side cue for words from the animacy task, *t*(47) = -5.99, *p* < 0.001, but not for the mobility task, *t*(47) = 0.62, *p* = 0.54. However, the nature of the *Group* x *Cue* interaction was less clear, as there were no significant group differences for accuracy under the Question or the Side cue when the data were collapsed across the two encoding tasks, *t*s < 1.20, *p*s > 0.23. Inspection of Figure 4A suggested that the cue effect varied by group for words from the mobility task, but not the animacy task. Consistent with this impression—and despite the lack of a three-way interaction (*F* < 1)—there was a significant *Group* x *Cue* effect for words from the mobility task, *F*(1, 46) = 6.45, *p* = 0.01 (left panel). This was not true for words from the animacy task, *F* < 1, where accuracy was characterized by a strong main effect of *Cue*, *F*(1, 46) = 35.89, *p* < 0.001 (right panel).

The nature of these results is highlighted in Figure 4B, which plots “Question minus Side” accuracy difference scores. For words from the mobility task (left panel), the Question minus Side difference score was greater than zero in the MDD group, *t*(23) = 2.06, *p* = 0.051, but not the controls, *t*(23) = -1.49, *p* = 0.15, and a between-groups *t*-tests revealed more positive difference scores in depressed adults, *t*(46) = 2.54, *p* = 0.015, *d*  = 0.73. By contrast, for words from the animacy task (right panel), the Question minus Side difference scores were significantly more negative than zero in both groups, *t*s < -4.25, *p*s < 0.001, and there was no group difference, *t*(46) = 1.00, *p* = 0.32, *d* = 0.29.

[PLEASE INSERT FIGURE 4 ABOUT HERE]

In summary, the cue effect varied by group for words from the mobility task: depressed adults showed better accuracy under the Question vs. the Side cue, but controls did not. By contrast, the cue effect was stable across groups for words from the animacy task, where accuracy was significantly lower accuracy under the Question vs. the Side cue.

**Retrieval: source confidence**. Figure 5A shows the confidence data. These were characterized by a main effect of *Task*, *F*(1, 46) = 6.72, *p* = 0.01, reflecting higher confidence in response to words from the mobility vs. animacy task, *t*(47) = 2.75, *p* = 0.008, *d* = 0.13, as well as a main effect of *Cue*, *F*(1, 46) = 26.10, *p* < 0.001, that was qualified by a *Group* x *Cue* interaction, *F*(1, 46) = 4.37, *p* = 0.04. The interaction emerged because there was no group difference in the percentage of high confidence responses under the Question cue, *t*(46) *<* 1, but the depressed adults were less confident than controls in response to the Side cue, *t*(46) = 2.25, *p* = 0.03, *d* = 0.65. Finally, we used exploratory *t*-tests to confirm that, in the MDD group, confidence was higher under Question vs. Side for both the animacy task, *t*(23) = 4.35, *p* < 0.001, *d* = 1.01, and the mobility task, *t*(23) = 5.46, *p* < 0.001, *d* = 1.10. Thus, in contrast to the guessing and accuracy data, the cue effect on confidence did not vary by encoding task in the MDD group.

[PLEASE INSERT FIGURE 5 ABOUT HERE]

**Retrieval: RT**. RT data are shown in Figure 5B. The only significant effect was an effect of *Cue*, with slower responses to Question vs. Side cues, *F*(1, 44) = 194.99, *p* < 0.001, *d* = 1.53. *Group* was not significant, *F* <1, nor were any interactions with *Group*, *F*s < 2.89, *p*s > 0.09.

**Overall behavioral summary**. At encoding, mobility judgments were made more slowly and less accurately than animacy judgments. This appears to have supported retrieval as both groups guessed less and were more confident when responding to words from the mobility task. Furthermore, participants responded more slowly, guessed less, and were more confident under the Question vs. the Side cue. Importantly, the combination of words from the mobility task presented under the Question cue yielded improved performance in the MDD group. Depressed participants guessed least frequently in this cell of the design, and source accuracy for words from the mobility task was characterized by a *Group* x *Cue* interaction: in depressed adults, but not controls, accuracy was better under the Question cue vs. the Side cue. This contrasted sharply with accuracy in response to words from the animacy task, where both groups were less accurate under the Question cue vs. the Side cue. It also differed from the pattern seen for confidence, where the MDD group was less confident than controls under the Side cue for both encoding tasks. In short, the mobility task and Question cue led to few guesses and confident responding in everyone, and the combination of these factors boosted accuracy in the MDD group.

**ERPs**

**Conceptual and perceptual retrieval, collapsed over encoding task**.To test our *a priori* hypothesis, we conducted between-group tests on “Question minus Odd/Even” and “Side minus Odd/Even” difference waves, collapsed across encoding task as in prior studies. We expected group differences in the former but not the latter contrast, but in fact we found no group differences in either contrast (smallest cluster *p* = 0.29). Thus, we present the results collapsed across groups in Figures 6 and 7.

[PLEASE INSERT FIGURE 6 AND TABLE 2 ABOUT HERE]

Figure 6 depicts the “Question minus Odd/Even” contrast and Table 2 lists electrodes where condition effects were observed. As shown in the top panel, from 400-800 ms there were two clusters of differential activity. As expected, Question hits elicited more positive ERPs than Odd/Even hits over left parietal electrodes, consistent with many prior studies of recollection (Rugg & Curran, 2007). In addition, there was a relative negativity for Question hits vs. Odd/Even hits over right frontal electrodes in this time window. As shown in the middle and bottom panels, later intervals (800-1400 ms and 1400-2000 ms) were characterized by sustained negative polarity potentials for Question hits relative to Odd/Even hits over left frontal and right occipital sites. The left frontal result replicates a study that linked this potential specifically to conceptual retrieval (Bergström et al., 2013), while the LPN over posterior sites is a consistent observation in many studies of both conceptual and perceptual source memory (Bergström et al., 2013; Johansson & Mecklinger, 2003; Mecklinger et al., 2007).

For the “Side minus Odd/Even” contrast (Figure 7, Table 3), no differences between the conditions were seen from 400-800 ms. However, strong condition effects were observed from 800-1400 ms and 1400-2000 ms. In these windows, a sustained LPN in response to Side hits was seen over the posterior midline, extending from anterior parietal to occipital sites.

[PLEASE INSERT FIGURE 7 AND TABLE 3 ABOUT HERE]

***Question* minus *Side*, animacy task**. Our second group-level analysis was designed to more closely follow the accuracy results. The retrieval cues strongly affected source accuracy for words from the animacy task, with worse performance under the Question vs. Side cue in both groups (Figure 4A, right panel). To probe the neural correlates of this effect, we computed “Question minus Side” difference waves for words from the animacy task. A between-groups test revealed no reliable differences (smallest cluster *p* = 0.20), thus we present data collapsed across groups in Figure 8 (also see Table 4). This contrast revealed a broadly distributed negativity that was focused over left fronto-central scalp from 400-800 ms, dispersed over bilateral fronto-central scalp from 800-1400 ms, and separated into left fronto-central and right centro-parietal clusters from 1400-2000 ms. Inspection of waveforms revealed a consistent pattern: relative to Side hits, Question hits elicited more negative potentials, with below-baseline activity especially evident from 1400-2000 ms.

[PLEASE INSERT FIGURE 8 AND TABLE 4 ABOUT HERE]

***Question* minus *Side*, mobility task**. Finally, we computed “Question minus Side” difference scores for words for the mobility task and compared responses across the two groups. As shown in Figure 9 (see also Table 5), this contrast was associated with group differences over left centro-parietal scalp between 400-800 ms (Figure 9, top) and 800-1400 ms (Figure 9, middle). In these intervals, the depressed and healthy groups generated similar responses for Question hits, but the depressed group showed a weaker response for Side hits. Indeed, follow-up *Group* x *Cue* ANOVAs on mean amplitudes averaged over all electrodes in these clusters (Figure 9, bottom) revealed *Group* x *Cue* interactions in both intervals (*F*s > 14, *p*s < 0.001). These reflected significant control > MDD effects for Side hits (*t*s > 2.2, *p*s < 0.025, *d*s > 0.67), but not Question hits (*t*s < 0.7, *p*s > 0.52, *d*s <0.19). Moreover, the MDD group generated stronger responses on Question vs. Side hits in both time windows (*t*s > 3.1, *p*s < 0.006, *d*s >0.54), but controls showed the opposite pattern—stronger responses to Side vs. Question hits (*t*s > 2.0, *ps* < 0.056, *ds* >0.20). This pattern of ERP effects is quite similar to the pattern seen for source accuracy (Figure 4).

[PLEASE INSERT FIGURE 9 AND TABLE 5 ABOUT HERE]

**Individual Differences**

To look for brain/behavior relationships across the groups, we first computed the mean amplitude of the left centro-parietal “Question minus Side” ERP difference waves for words from the mobility task from 400-800 ms and 800-1400 ms, averaging over the electrodes that showed group differences in the top and middle panels of Figure 9, respectively. The amplitudes of the ERPs in these intervals were highly correlated, *r* = 0.84, *p* < 0.001. These values were positively correlated with “Question minus Side” accuracy difference scores, although the relationships were modest and only significant in the later interval (400-800 ms, *r* = 0.18, *p* = 0.21; 800-1400 ms, *r* = 0.28, *p* = 0.05).

Next, we considered individual differences in the MDD group considered alone, again focusing on “Question minus Side” comparisons for words from the mobility task. First we computed Pearson correlations to examine associations between “Question minus Side” accuracy for words from the mobility task and depressive severity (BDI-II total), anhedonia (MASQ-AD), general distress associated with anxiety (MASQ-GDA), anxious arousal (MASQ-AA), brooding rumination (RRS-Brooding), or sleep disruption (PSQI total). We found negative relationships between accuracy and both general anxiety (Figure 10A; *r* = -0.42, *p* = 0.04) and anxious arousal (Figure 10B; *r* = -0.47, *p* = 0.02). No other result was significant, but there was a marginal relationship with BDI-II total score (*r* = -0.35, *p* = 0.09). To determine whether the relationships with anxiety were distinct from the relationship with BDI-II, we computed hierarchical regressions in which “Question minus Side”/*mobility* accuracy served as the criterion variable, BDI-II scores were added in step 1, and either MASQ-GDA or MASQ-AA were entered in step 2. Neither model was significantly improved by addition of the MASQ anxiety measures, and neither of the MASQ measures significantly predicted accuracy with BDI-II in the model (all *p*s > 0.11). This is not entirely surprising as BDI-II scores were strongly correlated with MASQ-GDA and MASQ-AA scores, *r*s > 0.68, *p*s < 0.001, suggesting that these self-report measures captured overlapping variance in psychological distress.

Next we examined relationships with “Question minus Side” ERP difference waves, computed for words from the mobility task. To maximize sensitivity, we did not restrict the analysis to sites that showed group differences (Figure 9). Instead, we extracted data from electrodes that showed condition effects in the MDD group considered alone (Figure 10C). These electrodes were predominantly located over left parietal scalp, and significant condition effects were evident in all three time windows (400-800 ms, 800-1400 ms, 1400-2000 ms). The mean ERP amplitude from these electrodes was correlated across the windows (*r*s > 0.39, *p*s < 0.053), but we found no relationship between source accuracy and ERP amplitude in any interval (|*r*|s < 0.33, *p*s > 0.12). Moreover, the ERP amplitudes were not related to the self-report measures, with one exception: we found negative relationships between PSQI scores and ERP amplitudes from 400-800 ms (Figure 10D; *r* = -0.48, *p* = 0.02) and 800-1400 ms (Figure 10E; *r* = -0.50, *p* < 0.02).

To confirm that these result did not simply reflect depressive severity, we computed hierarchical regressions with ERP amplitude as the criterion, entering BDI-II and PSQI scores in steps 1 and 2. PSQI predicted ERP amplitude after accounting for BDI-II in both intervals (400-800 ms; β = -0.45, *p* = 0.03; 800-1400 ms; β = -0.49, *p* = 0.03), and adding PSQI improved both models (Δ*R*2s > 0.16, Δ*F*s > 4.5, *p*s < 0.05). Thus, “Question minus Side”/*mobility* ERP amplitude was lowest in those depressed adults who reported chronic sleep disruption, and this effect did not simply reflect depressive severity.

[PLEASE INSERT FIGURE 10 ABOUT HERE]

**Discussion**

This study yielded three novel findings. First, there was a strong *Cue* x *Task* interaction for source accuracy as conceptual and perceptual retrieval were similarly accurate for words from the mobility task, but conceptual accuracy was significantly worse than perceptual accuracy for words from the animacy task. This dependency of conceptual source retrieval on the encoding task was unanticipated as prior imaging work with this paradigm has reported source accuracy collapsed over the different encoding tasks (Bergström et al., 2013; Simons, Gilbert, et al., 2005; Simons, Owen, et al., 2005). However, the result is consistent with a key finding from behavioral studies of multidimensional source memory: several contextual attributes (e.g., encoding task, encoding location) can be bound to a single item, and retrieval of any such attribute appears to be independent from retrieval of the others (Hicks & Starns, 2016; Starns & Hicks, 2005b; Vogt & Bröder, 2007). For instance, Starns and Hicks

Second, depressed people benefit from the mobility/Question combo. You can clearly see that in guessing and accuracy data, and it’s reflected in the left parietal ERPs. Seems more related to accuracy than confidence although we cannot make that argument strongly.

Third, accuracy and ERPs in depression appear to be related to the symptom severity and sleep disturbance, respectively. This is interesting because . . .

However, these conclusions must be tempered by some key limitations of the experimental design. Most importantly, no new items so it is really not possible to say whether what we’re seeing here is due to accuracy or bias. Also, we assume based on prior literature that the best place to see depression effects is on recollection, but it would be good to look at familiarity and we cannot do that here.

Also, how come MDDs did not do worse?

Do you want to interpret the Q v O/E and S v O/E maps at all?

You need to check out the ACSIM measure and the Vogt and co paper critiquing it, b/c S&H say that this measures is insensitive to bias to claim one source over the other.

Nonetheless, we think this is a particularly important topic worthy of future research, especially because impaired retrieval in depression has practical consequence (MEPS, inability to imagine future which may be key in therapy) and because it is increasingly targeted for treatment. Given these important efforts, a better understanding of the basic cognitive processes and neural systems affected by depression in the context of episodic retrieval is highly desirable.

This study yielded two sets of behavioral and ERP findings. Relative to controls, depressed adults were less accurate and less confident in their memories, and they showed reduced parietal ERP amplitude from 400-800 ms. The negative effect of MDD on memory was modest, but in addition to reporting lower confidence than controls in all four cells of the design, the depressed adults were numerically less accurate in three cells. Worse performance in 7/8 cells is improbable under the null (binomial test, *p* = 0.035 one-tailed), thus recollection and brain activity indexing recollection were weaker in MDD.

However, depressed adults showed excellent memory for words from the mobility task presented under the Question cue, which we interpret as reflecting sustained attention. At encoding, the mobility task elicited longer RTs and lower accuracy than the animacy task. It is easy to see why—for example, because trees sway in the breeze, deciding whether *oak* is “mobile” is harder than deciding whether an oak is alive—and we think the additional consideration needed to render mobility judgments led to deeper encoding. At retrieval, the Question cue elicited longer RTs and more confident responses than the Side cue, suggesting more extended and successful memory searches. Thus, directing conceptual retrieval at words from the mobility task pairs a deep retrieval search with deep encoding. Depressed adults can perform well under these conditions (P.T. Hertel & Rude, 1991; Paula T. Hertel & Hardin, 1990), and our ERP data highlight a candidate neural mechanism: sustained recruitment of left parietal cortex, which was otherwise hypoactive.

We speculate that this ERP effect may track relatively effortless recovery of episodic details, because left parietal activity was not observed when words from the animacy task were presented under the Question cue; instead lasting activity over left PFC was seen. Animacy decisions were made quickly at encoding, consistent with shallower encoding in this condition, and accuracy under the Question cue was lower following the animacy vs. mobility task (Figure 2A, right). Therefore, the left PFC activity may reflect additional cue elaboration needed to generate candidate memories following relatively poor encoding, or possibly post-retrieval monitoring or selection. Interestingly, the only other imaging study of source memory in MDD we know of reported increased left frontal activation during recollection attempts in depressed adults (Van Eijndhoven et al., 2013). That study did not manipulate encoding difficulty, but we predict that left PFC activation during source retrieval in MDD will be strongest when encoding is shallowest and recollection is weakest.

Our data may have treatment implications. As described earlier, imprecise retrieval is associated with depression and enhancing retrieval can speed recovery. Consideration of treatment mechanisms suggests an explanation. During cognitive behavioral therapy (CBT), patients recall difficult episodes from their lives and then reappraise them to reduce distress; here the importance of accurate retrieval is self-evident. But patients are also asked to imagine similar situations unfolding in the future so they can envision themselves effectively using new coping skills (Holmes, Arntz, & Smucker, 2007). Imagining future events depends on the same parieto-hippocampal circuitry that supports retrieval (Madore, Szpunar, Addis, & Schacter, 2016), and we have shown that activity in these circuits is blunted in MDD but can recover with adequate support. By extension, we speculate that effective CBT may be associated with improved functioning in parieto-hippocampal circuits. Given links between antidepressant effects and both functional and structural changes in the hippocampus (Santarelli et al., 2003), this argument may extend to psychopharmacological interventions as well. Finally, we expect that a sleep intervention would enhance memory retrieval in MDD, based on the negative relationship between sleep quality and ERP amplitudes observed in Figure 6.

In summary, this study provides novel insight into the impact of depression on brain activity during retrieval. The central role of parieto-hippocampal activity in episodic memory is already well-known. These data indicate that the same circuitry may play an important but underappreciated part in depression and its treatment.

**Acknowledgements**

The authors gratefully acknowledge Victoria Lawlor for assistance with recruitment and participant testing, and Dr. Diego Pizzagalli for helpful comments on a draft of the manuscript. Funding: the study was supported by NIMH grant R00 MH094438-03 (D.G.D) and by generous funding from McLean Hospital. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Financial Disclosures**

Dr. Dillon has received consulting fees from Pfizer, Inc., for work unrelated to this project. Ms. Barrick reports no biomedical financial interests or potential conflicts of interest.

**References**

**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*. Encoding (A) response accuracy and (B) correct response time (RT). Responses were slower and less accurate in the mobility task vs. the animacy task. There were no group differences, but data from controls and depressed participants are shown separately for comparison. Error bars = SEM.

*Figure 3*. Mean number of guesses by group, encoding task, and retrieval cue. Data are from 17 depressed and 11 healthy participants who guessed at least once in every condition. There was a main effect of *Task,* as all participants guessed less for words from the mobility (left panel) vs. the animacy task (right panel). There was also a main effect of *Cue*, as participants guessed less for words shown under the Question cue (blue bars) vs. the Side cue (green bars). In the MDD group, the cue effect on guessing was pronounced for words from the mobility task. Error bars = SEM, \**p* = 0.007.

*Figure 4*. Source memory accuracy (A) under Question (blue bars) and Side (green bars) cues, and for (B) Question minus Side difference scores. In both panels, the left column shows data for words from the mobility task, and the right panel shows data for words from the animacy task. There was a *Group* x *Cue* interaction for words from the mobility task, because the MDD group showed better memory under the Question vs. Side cue but the controls did not. By contrast, the animacy task was characterized by a main effect of *Cue*, as both groups show lower accuracy under Question vs. Side. Bar heights correspond to mean, error bars = SEM, \* *p* = 0.015.

*Figure 5.* (A) The percentage of high confidence responses as a function of group, cue, and task. All participants were more confident when responding to words from the mobility task (left panel) vs. the animacy task (right panel) and when responding to the Question cue (blue bars) vs. the Side cue (green bars). Depressed adults were less confident than controls in response to the Side cue but not the Question cue. Exploratory *t*-tests revealed thatdepressed adults were significantly more confident when responding to the Question vs. Side cue for words from both encoding tasks. (B) Mean correct RT data; all participants responded more slowly to the Question cue (blue bars) vs. the Side cue (green bars). Error bars = SEM, \**p*s < 0.001.

*Figure 6*. Mass univariate analysis of *Question* minus *Odd/Even* difference waves, from 400-800 ms (top), 800-1400 ms (middle), and 1400-2000 ms (bottom). The data are collapsed across groups as there were no significant between-group differences. On the topographies, electrodes in significant clusters are marked with white circles. The electrode that showed the strongest condition effect in each cluster is marked in red, and waveforms from that electrode are plotted separately for each condition, with the time window shaded in gray. Electrode numbers (e.g., “e109”) give the position on the EGI cap—see the Supplement for a complete map.

*Figure 7*. Mass univariate analysis of *Side* minus *Odd/Even* difference waves, from 400-800 ms (top), 800-1400 ms (middle), and 1400-2000 ms (bottom). The data are collapsed across groups as there were no significant between-group differences. On the topographies, electrodes in significant clusters are marked with white circles. The electrode that showed the strongest condition effect in each cluster is marked in red, and waveforms from that electrode are plotted separately for each condition, with the time window shaded in gray. Electrode numbers (e.g., “e83”) give the position on the EGI cap—see the Supplement for a complete map. No significant differences between the conditions were found from 400-800 ms.

*Figure 8*. Mass univariate analysis of *Question* minus *Side* difference waves, for words from the animacy task, from 400-800 ms (top), 800-1400 ms (middle), and 1400-2000 ms (bottom). The data are collapsed across groups as there were no significant between-group differences. On the topographies, electrodes in significant clusters are marked with white circles. The electrode that showed the strongest condition effect in each cluster is marked in red, and waveforms from that electrode are plotted separately for each condition, with the time window shaded in gray. Electrode numbers (e.g., “e16”) give the position on the EGI cap—see the Supplement for a complete map. An orange line separates the two clusters identified from 800-1400 ms.

*Figure 9*. Group differences in the mass univariate analysis of *Question* minus *Side* difference waves, for words from the mobility task, from 400-800 ms (top) and 800-1400 ms (middle); bottom panel shows the data averaged over all significant electrodes in each cluster, in each time window. On the topographies, electrodes in significant clusters are marked with white circles. The electrode that showed the strongest condition effect in each cluster window is marked in red, and waveforms from that electrode are plotted separately for each condition, with the relevant time window shaded in gray. Electrode numbers (e.g., “e16”) give the position on the EGI cap. The data in each time window were characterized by *Group* x *Cue* interactions: depressed and healthy adults generated similar responses to the Question cue, but controls generated a significantly stronger response to the Side cue.

*Figure 10*. In the MDD group, there was a negative correlation between “Question minus Side” accuracy difference scores for words from the mobility task and (A) general distress associated with anxiety (MASQ-GDA) and (B) anxious arousal (MASQ-AA). (C) Topographic maps of Question minus Side differences for words from the mobility task in the MDD group considered alone, with electrodes that showed significant condition effects highlighted in white. There were negative relationships between PSQI scores and Question minus Side ERP amplitudes from these electrodes in the (D) 400-800 ms and (E) 800-1400 ms time windows.