**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) affects episodic retrieval, but few functional imaging studies have examined the underlying neural mechanisms. Therefore, we collected event-related potentials (ERPs) from 24 unmedicated adults with MDD and 24 healthy controls during source memory retrieval. 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 seen 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, accuracy for words from the animacy task was characterized by a main effect of *Cue*:conceptual source accuracy was worse than perceptual source accuracy across groups, and ERP difference waves tracking this effect revealed 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 for its association with biased emotional memory. Relative to healthy controls, depressed adults 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 depression. In a meta-analysis of 22 neuropsychological studies, Zakzanis and colleagues (1998) found strong negative effects of depression on episodic memory tests conducted with neutral stimuli. Similarly, Airaksinen et al. (2004) used neutral words to demonstrate impaired free and cued recall in 187 depressed adults recruited from the community. Finally, a meta-analysis of over 30 years of work with the Cambridge Neuropsychological Test Automated Battery (CANTAB) indicated that, relative to healthy adults, depressed individuals show significantly impaired episodic memory, despite the fact that the CANTAB tasks use neutral stimuli (Rock, Roiser, Riedel, & Blackwell, 2014).

What underlies the memory impairment for non-emotional stimuli in depression? One obvious candidate is poor encoding. Consistent with this, 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. But a more nuanced result was presented by Zakzanis et al. (1998), who noted that the effect of depression on memory varied according to the encoding task: larger negative effects were seen for tasks that provided less support during learning (e.g., memorization of uncategorized vs. categorized word lists). The *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 strategies to enhance encoding but that, 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 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 the words, 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 presumably impair encoding, Hertel and Rude predicted a negative effect of depression specifically 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, depression impaired memory when encoding was unconstrained, but the provision of a task that engaged attention and encouraged elaborative encoding reduced the deficit.

The cognitive initiative framework also applies to retrieval. 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 invariably 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, effortful searching, and post-retrieval monitoring 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 identical words appeared on both tests, healthy participants were more likely than depressed adults to notice this regularity and exploit it to improve recognition (i.e., by referring back to the implicit test). However, once the relationship between the tests was highlighted, the group difference disappeared. Moreover, when participants were given a detailed strategy for improving their memory, depressed adults showed an especially large benefit. These results provide additional support for the claim that depression impairs episodic memory for neutral stimuli when the encoding and/or retrieval tasks are relatively unstructured, but that engaging tasks (or detailed instructions concerning optimal strategies) can rescue performance.

Given the extensive and elegant behavioral work on these issues, the paucity of data regarding relevant neurobiology is surprising. In particular, although there are many studies of hippocampal volumes in depression (for review, see G. MacQueen & Frodl, 2011) and some functional imaging investigations 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 data on memory retrieval in MDD. This does not reflect lack of desire—about 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. Because depression affects recollection more 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). Importantly, source memory depends heavily on recollection—although familiarity can play an important role (Mollison & Curran, 2012)—and there is evidence that it is disrupted in depression (Degl’Innocenti & Bäckman, 1999). To increase our chances of detecting an effect of MDD, we adapted a design that recruits neural systems engaged by both conceptual and 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 an animacy or mobility judgment. At test, they were cued to recall the presentation side (perceptual source) and encoding task (conceptual source). A recent fMRI/ERP study in healthy adults (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 also found 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).

Based on this prior work, we expected depression to selectively impair conceptual source memory, for two reasons. 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 reasoned that conceptual retrieval would be especially disrupted in MDD. Second, rumination is negatively associated with episodic memory in depression (Paula T. Hertel & T., 1998), and rumination activates PFC circuits that are particularly active during conceptual retrieval (Cooney, Joormann, Eugène, Dennis, & Gotlib, 2010). Thus, we expected depressive rumination to preferentially impair conceptual retrieval. Because prior work has presented the retrieval data collapsed over different encoding tasks, we did not expect the accuracy of conceptual or perceptual retrieval to vary as a function of 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 words, the mobility task was more difficult and elicited longer response times (RTs). This had consequences for retrieval. First, and consistent with the cognitive initiative framework, we found that when sustained, focused analysis at encoding was followed by conceptual retrieval, performance in the MDD group was surprisingly good. Thus, this experiment provided an opportunity to examine neural activity mediating accurate source memory in MDD when encoding and retrieval are well-supported. Second, we found that conceptual retrieval was strongly shaped by the encoding tasks but perceptual retrieval was not, and this was true in both groups. The discovery that retrieval of one contextual attribute (i.e., task) was more sensitive to the encoding manipulations than retrieval of another contextual attribute (i.e., location) is consistent with recent studies of multidimensional source memory (Hicks & Starns, 2016b; Starns & Hicks, 2005a; Vogt & Bröder, 2007), which have shown that retrieval of different contextual attributes occurs independently. Thus, this study also shows that in both healthy and depressed adults, brain activity during source retrieval is governed by interactions between retrieval cues and the encoding processes that were performed on particular items.

**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. They were screened by phone or online, at which time the Beck Depression Inventory II (BDI; Beck, Steer, & Brown, 1996) was administered. Prospective subjects 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 episodic memory (Plihal & Born, 1997) and because there is substantial evidence linking depression and other psychiatric disorders to sleep disruption (e.g., Deldin, Phillips, & Thomas, 2006; Wulff, Gatti, Wettstein, & Foster, 2010). Thus, we expected negative relationships between these measures on one hand and both memory accuracy and ERP indices of successful 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. WTAR data from four non-native English speakers (controls, *n* = 2; MDD, *n* = 2) were not analyzed, as this measure may not be valid in non-native speakers.

**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 significant 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?” (animacy task) or “mobile/immobile?” (mobility task). 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 16 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 to 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, 2005b), “guess” was included as a response option. Participants were asked to select this 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 *log*(RT) exceeded the participant’s mean±3SD (< 4% of encoding trials, < 1% 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. ANOVAs were computed on *log*(RT) data because of improved fit to normality (raw RT data were positively skewed), but the figures and descriptive statistics use untransformed RT data.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 intended to encompass 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 is appropriate for detecting neighboring electrodes in regions where they 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 group difference (Bullmore et al., 1999). This null distribution was used to estimate the probability of observing the clusters actually detected when the 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 accuracy, confidence, and left parietal ERPs associated with recollection. Within the depressed group, we examined relationships between source accuracy, confidence, 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) = 13.03, *p* < 0.001, 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) = 50.20, *p* < 0.001, with slower responses for mobility vs. animacy judgments. The RT analysis also revealed a *Group* x *Task* x *Side* interaction, *F*(1, 44) = 8.02, *p* = 0.007, but separate ANOVAs for words presented on the left and right did not reveal significant *Group* x *Task* interactions, *F*s < 3.76, *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 encoding accuracy and slower RTs, but there were no reliable group differences.

[PLEASE INSERT FIGURE 2 ABOUT HERE]

**Retrieval: Odd/Even trials**. There were also no group differences on Odd/Even trials, which were characterized by extremely high accuracy (controls: 98.44±3.92%; MDD: 99.13±1.37%; *t*(46) = -0.82, *p* = 0.42), fast correct RTs (controls: 867.72±301.94 ms; MDD: 781.28±227.83 ms; *t*(44) = 1.10, *p* = 0.28), and highly confident responses (percentage of high confidence trials: controls: 99.61±0.68%; MDD: 99.87±0.35%; *t*(46) = -1.68, *p* = 0.10).

**Retrieval: guessing**. Five participants (three controls, two MDD) never guessed and a further 12 participants (8 controls, 4 MDD) did not guess in at least one cell of the design. Data from the remaining participants (13 controls, 18 MDD) are shown in Figure 3 and were included in the *Group* x *Cue* x *Task* ANOVA; because only a single guess was recorded on an Odd/Even trial, that condition was omitted. The ANOVA yielded an effect of *Cue*, *F*(1, 29) = 7.75, *p* = 0.009, reflecting fewer guesses under the Question vs. Side cue, and an effect of *Task*, *F*(1, 29) = 15.98, *p* < 0.001, reflecting fewer guesses following 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, but there was a trend for a *Group* x *Cue* x *Task* interaction, *F*(1, 29) = 3.62, *p* = 0.07. Inspection of Figure 3 suggested that this was due to a pronounced cue effect for words from the mobility task in the MDD group. Indeed, a *t*-test confirmed that, for words from the mobility task, depressed participants guessed less frequently under the Question vs. the Side cue, *t*(17) = -3.29, *p* = 0.004, *d* = 0.77; the cue effect was not reliable for words from the animacy task, *t*(17) < 1. In controls there was no reliable cue effect for either encoding task considered alone (*t*s < 1.33, *p*s > 0.20). 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*(17) = -3.74, *p* = 0.002, *d* = 0.82, or words from the animacy task shown under the Side cue, *t*(17) = -3.59, *p* = 0.002, *d* = 0.85. In sum, all participants guessed less to words from the mobility vs. animacy task and in response to the Question vs. Side cue. In the MDD group, words from the mobility task presented under the Question cue elicited especially few guesses.

**Retrieval: source accuracy**. The source accuracy data are shown in Figure 4A; the left and right panels show accuracy for words from the mobility and animacy tasks, respectively. These were characterized by main effects of *Cue*, *F*(1, 46) = 20.33, *p* < 0.001, and *Task*, *F*(1, 46) = 30.43, *p* < 0.001, which were qualified by two interactions: *Cue* x *Task*, *F*(1, 46) = 21.69, *p* < 0.001, and *Group* x *Cue*, *F*(1, 46) = 5.49, *p* = 0.02. 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) = -6.08, *p* < 0.001, but not for the mobility task, *t*(47) = 0.25, *p* = 0.80 (an alternative way to view this interaction is to note that accuracy under the Question cue was significantly better for words from the mobility task vs. the animacy task, t(47) = 5.65, *p* < 0.0001, whereas accuracy under the Side cue did not differ by encoding task, *t*(47) = 1.89, *p* = 0.064).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 Side cue when the data were collapsed across the encoding tasks, *t*s < 1.16, *p*s > 0.25. 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) = 5.92, *p* = 0.02. 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) = 36.81, *p* < 0.001.

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 difference score was marginally greater than zero in the MDD group, *t*(23) = 1.77, *p* = 0.090, but not in the controls, *t*(23) = -1.67, *p* = 0.11, and a between-groups *t*-test was significant, *t*(46) = 2.43, *p* = 0.019, *d*  = 0.70. By contrast, for words from the animacy task (right panel), the difference scores were more negative than zero in both groups, *t*s < -4.29, *p*s < 0.001, and there was no group difference, *t*(46) = 0.92, *p* = 0.36, *d* = 0.27. For both groups, Question minus Side difference scores were more positive for the mobility vs. the animacy task, *t*s > 2.7, *ps* < 0.02.

[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 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) = 7.91, *p* = 0.007, reflecting higher confidence in response to words from the mobility vs. animacy task, *t*(47) = 2.94, *p* = 0.005, *d* = 0.13, as well as a main effect of *Cue*, *F*(1, 46) = 24.85, *p* < 0.001, that was qualified by a *Group* x *Cue* interaction, *F*(1, 46) = 4.24, *p* = 0.05. The interaction emerged because there was no group difference in the percentage of high confidence responses under the Question cue, *t*(46) *<* 1, but depressed adults were less confident than controls under the Side cue, *t*(46) = 2.25, *p* = 0.03, *d* = 0.65. We computed exploratory *t*-tests to confirm that, in the MDD group, confidence was higher under Question vs. Side for both the mobility task, *t*(23) = 5.35, *p* < 0.001, *d* = 1.10, and the animacy task, *t*(23) = 4.36, *p* < 0.001, *d* = 1.01. 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. There was only a significant effect of *Cue*, with slower responses to Question vs. Side cues, *F*(1, 44) = 267.92, *p* < 0.001, *d* = 1.63. *Group* was not significant, *F* <1, nor were any interactions with *Group*, *F*s < 1.

**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, words from the mobility task presented under the Question cue elicited good performance in the MDD group. Depressed participants guessed least frequently in this cell of the design, and 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 vs. the Side cue. These effects 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, (1) the mobility task and Question cue led to few guesses and confident responding in all participants, (2) the combination of these factors boosted accuracy in the MDD group, and (3) accuracy was lower under the Question vs. the Side cue for all participants when responding to words from the animacy task but not the mobility task. The observation of significantly poorer retrieval under the Question vs. Side cue for words from the animacy but not mobility task (*Cue* x *Task* interaction) demonstrates that conceptual retrieval was more strongly affected by the encoding tasks than perceptual retrieval.

**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 at all (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 reliable differences between the conditions were seen from 400-800 ms. However, strong 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. Robust activation over the posterior midline during retrieval of spatial information is consistent with both prior findings from this task (Bergström et al., 2013) and the larger “PMAT” framework (Ritchey, Libby, & Ranganath, 2015), which describes consistent activation of posterior midline (PM) as opposed to anterior temporal (AT) structures during retrieval of contextual information in general and spatial information in particular. Thus, both contrasts in our *a priori* plan yielded sensible results that were in line with prior findings, but neither provided insight into MDD.

[PLEASE INSERT FIGURE 7 AND TABLE 3 ABOUT HERE]

***Question* minus *Side*, animacy task**. Our second group-level analysis was designed to more closely track 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 stable 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 higher amplitude ERPs for 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 is similar to the pattern seen for 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 also positively correlated with “Question minus Side” accuracy and confidence difference scores, although the relationships were modest (**400-800 ms**: accuracy, *r* = 0.18, *p* = 0.21; confidence, *r =* 0.27, *p* = 0.06; **800-1400 ms**: accuracy, *r* = 0.29, *p* < 0.05; confidence, *r* = 0.28, *p* = 0.05). The accuracy and confidence difference scores were themselves correlated, *r* = 0.40, *p* = 0.004.

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 and depressive severity (BDI-II total), anhedonia (MASQ-AD), general distress associated with anxiety (MASQ-GDA), anxious arousal (MASQ-AA), brooding rumination (RRS-Brooding), and sleep disruption (PSQI total). We found negative relationships between accuracy and both general distress due to anxiety (Figure 10A; *r* = -0.42, *p* = 0.04) and anxious arousal (Figure 10B; *r* = -0.48, *p* < 0.02). No other result was significant, but there was a weak relationship with BDI-II total score (*r* = -0.32, *p* = 0.12). To determine whether the relationships with anxiety remained significant after accounting for depressive severity, we computed hierarchical regressions in which “Question minus Side” accuracy for words from the mobility task 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 MASQ measure predicted accuracy with BDI-II in the model, although the effect of MASQ-AA was marginally significant (*p* = 0.07). This was not entirely surprising as BDI-II, MASQ-GDA, and MASQ-AA scores were correlated, *r*s > 0.68, *p*s < 0.001, suggesting that they captured overlapping variance in psychological distress. Thus, we were not able to identify an effect of anxiety that was separable from greater depressive severity.

Next we examined relationships with “Question minus Side” ERP difference waves, computed for words from the mobility task, in the depressed group. 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 clusters were found in all three time windows (400-800 ms, 800-1400 ms, 1400-2000 ms). The mean “Question minus Side” ERP amplitude from these electrodes was correlated across the windows (*r*s > 0.39, *p*s < 0.053), but we found only a marginal relationship between source accuracy and ERP amplitude from 1400-2000 ms, *r* = 0.39, *p* < 0.06; no other relationships with accuracy or confidence approached significance (|*r*|s < 0.26, *p*s > 0.22). Moreover, ERP amplitudes were not related to the self-report measures with one exception: we found negative correlations 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 results 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* = 0.04; 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” ERP amplitude over left parietal scalp, in response to words from the mobility task, was lowest in depressed adults who reported chronic sleep disruption, and this did not simply reflect greater 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 across participants, as conceptual and perceptual retrieval were similarly effective for words from the mobility task, but conceptual accuracy was substantially worse than perceptual accuracy for words from the animacy task. In other words, conceptual memory was more strongly shaped by the encoding tasks than was perceptual memory. The dependency of conceptual retrieval on encoding was unanticipated as prior work with this paradigm has collapsed over the encoding tasks (Bergström et al., 2013; Simons, Gilbert, et al., 2005; Simons, Owen, et al., 2005). However, the results are similar to a key result from behavioral studies of multidimensional source memory: retrieval of any one contextual attribute of an item appears to occur independent of retrieval of any other attributes (Hicks & Starns, 2016a; Starns & Hicks, 2005a; Vogt & Bröder, 2007).

For instance, across six studies Starns and Hicks manipulated two dimensions of the encoding context, namely font size and word location (Starns & Hicks, 2005a) or the gender of faces (on which words were superimposed) and word location (Hicks & Starns, 2016b). They then provided cues relevant to one dimension at retrieval—for example, reinstating the font size of some words at test. If the various attributes of the encoded words were directly bound to each another, one would expect these cues to improve not only “within-dimension” retrieval (e.g., memory for font size), but also “across-dimension” retrieval (e.g., memory for word location). However, the six experiments yielded no evidence for across-dimension cuing, with the posterior distribution for this effect squarely centered on zero (see also Vogt & Bröder, 2007). Our demonstration that conceptual retrieval was more sensitive to the encoding tasks than perceptual retrieval provides additional evidence for the hypothesis that retrieval of one attribute (here, encoding task) can leave retrieval of other attributes (here, encoding location) unaffected (for additonal evidence pertinent to encoding manipulations, see Marsh, Hicks, & Cook, 2004).

The fact that attributes of an item can apparently be retrieved independently of course does not explain why conceptual retrieval was particularly sensitive to the encoding tasks in this experiment. We offer two possible mechanisms. First, it seems plausible that conceptual encoding was more effective for the mobility vs. the animacy task. Encoding RT was slower and encoding accuracy was lower on mobility (vs. animacy) trials, and because both tasks required analysis of the semantic properties of the words, we regard this as evidence of sustained analysis—deeper processing—in the mobility task. This inference is similar to one made by Dobbins and Wagner (2005), who found slower encoding RTs for pleasantness judgments vs. animacy judgments. They suggested that the basis for making pleasantness judgments varies more from item to item than does the basis for making animacy judgments, and thus concluded that the former involved more “sustained conceptual analysis” than the latter. Similarly, the cognitive processes involved in mobility judgments are likely more numerous and varied than those involved in animacy judgments—for example, it seems easier to decide whether or not *“*elm” refers to a living thing than to decide whether or not an “elm” is mobile, as trees sway in the breeze but clearly do not move on their own power. Consequently, Question accuracy may have been worse for words from the animacy task simply because participants generated relatively weaker encoding traces on animacy trials. The second possibility is closely related: perhaps the similarity of the semantic processing engaged by both encoding tasks, combined with deeper processing on mobility trials, resulted in heightened interference when words from the animacy task were presented under the Question cue; participants may have inadvertently retrieved material that seemed consistent with the processes engaged by the mobility task, and thus erroneously endorsed that conceptual source. Additional work is needed to determine whether either of these psychological accounts is accurate.

Whatever cognitive mechanisms were engaged, the results in Figure 8 suggest that their neural underpinnings differ from those typically implicated in episodic retrieval. Specifically, the “Question minus Side” contrast computed on words from the animacy task did not reveal positive ERPs over the left parietal cortex from 400-800 ms (seen in the “Question minus Odd/Even” [Figure 6] and “Question minus Side”/*mobility* [Figure 9] contrasts), nor did it reveal an LPN centered over posterior regions (as seen in the “Side minus Odd/Even” [Figure 7] and “Question minus Side”/*mobility* [Figure 9] contrasts). Instead, there was a sustained LPN that was broadly distributed but that appeared to have a fronto-central focus. To our knowledge, this specific ERP effect has not been previously reported in studies of episodic memory and its functional correlates are unclear. However, mediofrontal negativities are elicited by outcomes that are worse than expected, even when those outcomes do not demand a response, and it has been suggested that this reflects a signal to increase cognitive control in order to improve performance and the accuracy of predictions (Potts, Martin, Burton, & Montague, 2006). The tasks typically used to elicit mediofrontal negativities differ in many respects from our tasks, but the fact that Question/*animacy* accuracy was relatively poor lends plausibility to the idea that these ERPs may also reflect signals related to a need for additional cognitive control (i.e., to initiate a more extensive or efficient retrieval search in service of more accurate responding).

Second, we found surprisingly good performance—few guesses, high accuracy—in the MDD group when words from the mobility task were presented under the Question cue. We view the accuracy boost on Question/*mobility* trials in MDD as consistent with the cognitive initiative framework, because this cell of the design likely offered the most support at encoding and retrieval. As already noted, the mobility task appears to have elicited deeper processing than the animacy task. At retrieval, responses to the Question cue were made more confidently and markedly more slowly than responses to the Side cue, and therefore—although the link to “structure” is admittedly somewhat underdeveloped—we conclude that responses to the Question cue were made more deliberately and with a greater sense of efficacy than responses to the Side cue. Thus, the results shown in the left panels of Figure 4A and Figure 4B are in line with prior findings: given engaging encoding and retrieval tasks, episodic memory in depressed adults can be surprisingly good. Perhaps the most interesting and novel result from a clinical perspective is the fact that this behavioral result was linked to the amplitude of left parietal ERPs consistently associated with recollection (Rugg & Curran, 2007), as shown in Figure 9. Consequently, we suggest that whether a given episodic memory task yields better vs. worse performance in depressed vs. healthy adults may hinge on the extent to which this left parietal circuitry—and the core recollection network more broadly (Rugg & Vilberg, 2013)—is engaged.

Third, in the depressed group the amplitude of the left parietal ERPs was negatively correlated with self-reported sleep disruption over the last month. **This is . . . also bring in the link with BDI and MASQ here.**

These conclusions must be tempered by important limitations associated with the experimental design. First, we cannot conclusively link the ERP results in Figure 9 to accuracy vs. confidence. We are inclined to interpret them as reflecting changes in retrieval accuracy because the patterns in Figure 9 (bottom panel) and Figure 4 (top left panel) are quite similar, but across the groups the ERPs were weakly correlated with both accuracy and confidence. Furthermore, “Question minus Side” accuracy and confidence difference scores were significantly positively correlated for words from the mobility task, which suggests that cleanly dissociating effects associated with accuracy vs. confidence may not be possible in this dataset.

Second, and perhaps more importantly, we cannot be certain that the *Cue* x *Task* accuracy interaction truly reflects a difference in memory as opposed to response bias. That is, it is possible that participants were simply inclined to respond “mobility task” under the Question whenever they were uncertain, and this—rather than a difference in veridical recall—is what led to the interaction. Unfortunately, we know of no way to distinguish between these possibilities. It is possible to compute *d*’ for responses to the Question and Side cue in this task by arbitrarily calling correct responses to one source (mobility task, right side) “hits” and incorrect endorsements of these sources (i.e., responding “mobility task” to words from the animacy task, or “right” to words presented on the left) “false alarms” (Hicks & Starns, 2016b; Slotnick & Dodson, 2005). We have done this, and we find a reliable *Group* x *Cue* interaction that reflects (non-significantly) higher *d*’ scores for depressed vs. healthy adults under the Question cue and vice versa under the Side cue (analysis not shown). At the very least, these results indicate that our data are not confounded with a strong group difference in response bias, but the *d’* analysis is limited as the need to incorporate both encoding tasks into the calculation makes it impossible to detect the *Cue* x *Task* interactions that are clearly important here. Thus, better separating changes in accuracy from shifts in response bias is another goal for future work.

Third, the lack of new items in this paradigm is a mixed blessing. On the one hand, because all the words are “old” at retrieval, there is no reason to suspect that source decisions are confounded with gross differences in item memory, which is frequently the case in designs that require old/new judgments before source decisions are rendered (Murnane & Bayen, 1996). However, the lack of new items contributes to the aforementioned problem associated with teasing apart accuracy vs. response bias, and it also makes it impossible to determine whether familiarity made any contribution to performance or was affected by MDD. We were aware of this issue at the outset of the study and decided that it was not of critical importance given that MDD affects recollection much more strongly than familiarity (Paula T. Hertel & Milan, 1994; G. M. MacQueen et al., 2002), but it would be preferable to include new items in future studies.

We believe that this line of work is valuable despite the limitations associated with this study. Much of the early work on memory in depression was conducted by neuropsychologists who wished to know how to discriminate dementia from depression, or so-called “pseudo-dementia” (Burt et al., 1995). But memory deficits in depression deserve to be well-understood for their own sake, for several reasons. First, the hopelessness that characterizes depression is bound up in negative autobiographical memories, and thus a better understanding of the way in which these memories are encoded, consolidated, and retrieved could yield insight into the pathophysiology of MDD. Second, it is clear that autobiographical memory retrieval in depression is overly general (J. M. G. Williams et al., 2007), and this is linked to both poor problem-solving in the present and imprecise thinking about the future (J. M. Williams et al., 1996). Because cognitive behavioral therapy (CBT) depends on the patient’s ability to recall difficult episodes from their lives and then imagine similar situations unfolding more positively in the future (Holmes, Arntz, & Smucker, 2007), there are likely important but underappreciated links between episodic retrieval, prospection, and the ability to benefit from CBT. Indeed, the new field of “memory therapeutics” is finding that enhancing episodic memory—for non-emotional as well as emotional stimuli—leads to meaningful clinical improvements (for review, see Dalgleish & Werner-Seidler, 2014), and it is intriguing that imagining future events depends on parietal circuitry (Madore, Szpunar, Addis, & Schacter, 2016) very similar to what we have highlighted in Figure 9. Third, in addition to being debilitating, depression is often recurrent, making it a leading cause of years lived with disability worldwide (Ferrari et al., 2013) that costs the U.S. billions of dollars each year (Greenberg, Fournier, Sisitsky, Pike, & Kessler, 2015). There is an acute need for better treatments (Fournier et al., 2010), and there is evidence from non-human animals linking the effects of antidepressants to structural changes in brain regions critically implicated in episodic memory (Perera et al., 2007; Santarelli et al., 2003, 2008). Determining whether or not these findings extend to humans depends on a better understanding of how depression affects the neural machinery that supports episodic memory. The current study is a small step in this important direction.

**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.