1. Our results clearly show that the encoding task strongly affected participants’ ability to retrieve conceptual information. We now discuss this point and the limitation associated with the lack of new items in the manuscript, on pages X and Y, respectively.

2. First, we have taken the reviewer’s suggestion and added descriptive and statistical data on confidence (as well as accuracy and RT) for the parity judgment on Odd/Even trials (see page X).

3. As shown in Figure 5, the MDD group was less confident than the controls in response to the Side but not the Question cue, but this was not specific to either encoding task. This does not mirror the accuracy data, which showed a *Group* x *Cue* interaction for words from the mobility task, and which did not show a significant controls > MDD effect under the Side cue (or the Question cue). Thus, the reviewer’s point about lower confidence not tracking lower accuracy is well-taken. Consequently, on page X we now offer a more careful discussion of the confidence data.

4. As precedent, we note that when Dobbins & Wagner (2005) found slower encoding RTs for pleasant/unpleasant judgments than for animacy judgments, they suggested “that comparison of the former to the latter encoding trials would identify regions differentially engaged during sustained conceptual analysis” (p. 1773). Here, we are similarly arguing that the longer RTs observed for mobility versus animacy judgments implies “sustained conceptual analysis” (i.e., deeper processing) for the former relative to the latter trials. We have attempted to clarify this point on page X.

5. *The authors indicate that “observed ERP differences over parietal areas meant that ‘recollection was strongest under the Question cue and reduced in MDD’”. This is problematic because: (1) only correct responses were analyzed; (2) “I thought the behavioral data did not show an accuracy deficit for the MDD participants and thus claims that the ERPs correspond to group differences in the quality of memory evidence seem strained”; and (3) there is no evidence that individual differences in ERP amplitude correspond to individual differences in source accuracy or confidence.*

We appreciate the reviewer’s comments and have re-phrased the problematic passages—please see the new text about X on page Y in the Discussion.

6. Third, in the revised manuscript (page X) we note that there were positive relationships between left parietal Question minus Side ERP difference waves in the 400-800 ms and 800-1400 ms time windows, on the one hand, and Question minus Side source accuracy and confidence difference scores on the other.

7. We also added a note about the exploratory nature of the subsequent ERP analyses on page X of the Discussion.

8. As noted in response to point 2 (above), the absence of new items is a clear limitation, and we now discuss that on page X.

9. Finally, we note that prior studies of multidimensional source monitoring tasks also offer the “guess” response option (e.g., Starns & Hicks, 2005), and we have added references to those studies on page X.

10. *“The sentence ‘In short, memory retrieval is impaired in depression and enhancing it can bring lasting relief’ is unsubstantiated and needs clarification and references*.

We thank the reader for their careful reading of the manuscript. In response to this critique, we have edited the sentence (p. X) by adding references to work by Mark Williams, Tim Dalgleish, (and anyone else?), Filip Raes, and their colleagues. This work touches on the two topics mentioned in the sentence. First, several studies (i.e., X, Y, Z) provide evidence of retrieval deficits in depression, primarily in the context of autobiographical memory.

11. Instead, it seems that enhancing the specificity of memory retrieval may involve enhancements in executive control that have broad benefits on mood and cognitive function in depression. We have attempted to clarify these points in the manuscript (see page X), and we thank the reviewer for highlighting this need.

12. In the revision we devote more space to discussing about the lack of stronger negative effects of depression on memory accuracy (see page X). Speaking directly to the reviewer’s point, in the same place we specifically suggest that future neuroscientific studies of retrieval use emotional stimuli to see the expected memory deficit in MDD. In fact, such studies are underway in our laboratory.

13. As described in the manuscript (page X), we used a screening procedure to identify individuals who were in the midst of a Major Depressive Episode and who had a BDI-II score of at least 14 (the published cut-off for minimal depression).

14. It was on the basis of this statement that we adopted a criterion of 50% contaminated trials for rejection, and we have added a reference to Luck’s book on page X to indicate this.

15. We felt that loss of up to ~15% of the data (i.e., 18 of 128 channels) was a reasonable threshold for rejection. We now explain this on page X of the revised manuscript.

16. On page X of the revision, we now indicate that the first mechanism—additional cue elaboration—is more likely, based on a recent functional magnetic resonance imaging fMRI paper by Han and colleagues.

17. As noted in our response to Reviewer 5’s fifth critique (see above), we now put greater emphasis on the link between relatively intact source memory and increased left parietal ERPs in the Question/mobility condition for the MDD group (see page X)—this seems to be consistent with what the reviewer is driving.

18. We appreciate this comment and can see how early introduction of neural systems adds complexity to an already complicated discussion of the psychological constructs that mediate encoding and retrieval. In the revision, we have tried to take a more sequential approach of the kind the reviewer suggests—please see page X.

19. Therefore, seeing accuracy under the Question cue vary so strongly by encoding task was striking. We regret superficial treatment of this result, and in the revised manuscript it is now clearly emphasized (see pages X and Y).

20. As described on page X, we found that “Question minus Side”/*mobility* difference scores for accuracy and confidence were significantly correlated across the groups, *r*(46) = 0.42, *p* = 0.002, indicating that it may be difficult to tease apart these two factors.

21. Putting these two analyses together, we have advanced a cautious argument in the revision (see page X). Specifically, we indicate that the ERPs track accuracy more closely than confidence, but we also note that the correlations link the ERPs to both accuracy and confidence. A more definitive separation of accuracy and confidence in MDD will have to accomplished in a follow-up study better designed to tease apart these two factors.

22. Thus, we cannot advance a strong argument but we have added a reference to mental imagery on page X to reflect the reviewer’s interest, which we share.

23. A major shortcoming of the original manuscript was that our discussion of the PSQI data was very truncated due to space limitations. In the revision we devote more attention to this result (see page X).

24. Thus, we interpret the reviewer as suggesting that more drowsy depressed participants unhelpfully generate high amplitude alpha rhythms over the left parietal cortex, which might explain the negative correlation between chronic sleep disruption (as measured by PSQI) and the amplitude of the Question minus Side ERP effects (for words from the mobility task) seen over left parietal scalp. This is a fascinating suggestion, and we briefly acknowledge it on page X in the revision.

25. Thank you for this careful reading—we avoid referring to ERPs as indexing “activation” in the revised manuscript.

26. *On page 13, ‘no significant effects were seen in any time window’ is ambiguous, in the Abstract, ‘slasting’ is a typo, and on page 4, ‘loses’ is a typo*. Make these corrections or cut the relevant sentences.

27. We did this and found significant *Group* x *Cue* x *Task* interactions for 400-800 ms, *F*(1, 46) = 8.23, *p* = 0.006, and for 800-1400 ms, *F*(1, 46) = 5.09, *p* = 0.03. As described in the manuscript on page X, in both cases the *Group* x *Cue* interaction was significant for words from the mobility task, *Fs* > 14.3, *p*s < 0.0005.

28. When we considered the data across both groups, we found modest correlations between source accuracy, confidence, and left parietal ERP amplitudes for the “Question minus Side” contrast for words from the mobility task (**400-800 ms**; accuracy, *r* = 0.18, *p* = 0.21; confidence, *r* = 0.27, *p* = 0.06; **800-1400 ms**; accuracy, *r* = 0.28, *p* = 0.05; confidence, *r* = 0.28, *p* = 0.05), as described on page X. However, when we restricted these correlations to the MDD group, nothing approached significance, all *ps* > 0.12 (see page X).

29. Indeed, when we ran stepwise regressions predicting “Question minus Side”/*mobility* accuracy with BDI-II entered in Step 1 and either MASQ-GDA or MASQ-AA entered in Step 2, neither regression yielded a significant effect for the anxiety measures (χ2 < 2.7, *p*s > 0.11). This is not terribly surprising, as these three self-report measures were highly correlated (*rs* > 0.68, *p*s < 0.0002). We describe these new analyses in the manuscript on page X; although they do not support an argument for a selective effect of anxiety on memory, they certainly indicate that anxiety should be investigated in future studies of memory in psychopathology.

30. Unfortunately, space limitations made it difficult for us to describe our rationale for including the PSQI in the original manuscript. We do so in the revision on page X.

31. *In the Introduction, the authors hint at the Group* x *Cue* x *Task interaction for accuracy but do not explicitly describe it until later—it is worth stating it clearly here as well*.

Thank you for this suggestion; we now clarify the nature of the key interactions in the Introduction, on page X.

32. *On page 12 of the Results, the statement that “These data suggest that recollection was strongest under the Question cue and reduced in MDD*” *combines two main effects and reads as though there was an interaction, which there is not*.

The reviewer is correct and we apologize for the lack of clarity. This passage has been revised, please see page X.

33. Regarding **e**, we now report on correlations between source accuracy and confidence—please see page X.

34. For instance, if the reviewer compares the ERP bar graph at the bottom of Figure 9 with the accuracy data in Figure 4A, he or she will see the resemblance: in the MDD group there is a Question > Side effect, but in controls the pattern is reversed (Question < Side). By contrast, the confidence data in Figure 5A do not show such a pattern—for confidence, the pattern is Question > Side for both groups. Clearly this is not rock solid evidence and we do not wish to make too much of it, but on page X we point to this exercise as our reason for slightly favoring an accuracy account of the ERP effects.

35. Regarding point **f**, the reviewer’s comment (as well as a similar comment from Reviewer 5) made it clear that we did not plainly articulate our idea about RT and depth of encoding. We describe our thinking in more detail on page X of the revision.

36. In the revision, we devote significantly more space to the topic of rote memory difficulty in depression (see page X), and strive to be clear about the precise aspect of memory we are discussing (e.g., see line X on page Y).

37. *The manuscript does not carefully distinguish between recall and recognition within the broader concept of cued/uncued retrieval*.

We apologize for this and now provide important background information on page X.

38. On page X of the revision, we devote space to thinking about why the MDD group performed as well as it did.

On the nature of (non-emotional) memory impairments in depression

1. **Burt et al. (1995)**

* Meta-analysis of 99 recall and 48 recognition studies
* In their Table 2, they report extremely strong negative relationships between memory and free recall (Cohen’s *d* = 0.56, fail-safe Ns > 300 depending on the number of variables included in the study), with especially strong relationships for visual (as opposed to verbal) stimuli and for immediate (versus delayed) tests. There are many fewer studies of recognition memory that examine discriminability (*n* = 9 for “one-level” studies, as opposed to 54 such studies for recall), but these also support a negative relationship with depression, with *d* = 0.33 and fail-safe *N*s > 11. Regarding patient characteristics (Table 2), the effects are stronger with younger, unmedicated, inpatients. In terms of manipulations (Table 5), there is a stronger effect of depression on positive vs. neutral vs. negative memory.
* For recognition, some of this changes . . . although many studies only looked at hit rate, not discriminability. But with recognition you see stronger effects of depression on verbal vs. visual stimuli, and after a delay relative to immediate test. Again, the effects are stronger on younger inpatients relative to older outpatients.
* They make the point that many of these effects are not specific to depression and may be associated with “one or more overarching factors common to various forms of psychopathology (e.g., severity of illness, motivational deficits, effortful processing deficits)”

2. **Zakzanis et al. (1998)**

* Meta-analysis of 22 neuropsychological studies encompassing 726 adults with unipolar depression and 795 healthy controls. A nice aspect of this review is that it includes multiple measures of memory and other cognitive functions as well, so you have a shot at specificity.
* They find a median effect size of 0.52 (patients < controls), with several measures of episodic memory (and attention) showing effects greater than the median, but measures of semantic memory, primary memory, working memory, reasoning, and motor speed yielding weaker group differences (i.e., effect sizes below the median).
* They also find near-complete separation between the groups (only 10% overlap b/w distributions) on tasks characterized as “effortful”, versus substantial overall (73%) for tasks characterized as demanding only “superficial” processing.
* “depression had the largest effects on tests of episodic, declarative memory”
* Interestingly, they note that the negative effect of depression on declarative memory varies according to the amount of support provided at encoding—for instance, the California Verbal Learning Test, which uses lists of words sorted into categories, yielded weaker effects than the Ray Auditory Verbal Learning Test, which involves encoding 16 unrelated words. They regard this and similar evidence as consistent with an encoding deficit in depression than can be counteracted by external support (e.g., provision of categories that can be used to sort the lists and relate items to one another).

3. **Rock et al. (2014)**

* This is a meta-analysis of studies over 30+ years (1980-2012) that used the Cambridge Automated Neuropsychological Test Automated Battery (CANTAB), which is desirable as it ensures that everyone had the same tests, administered in the same way.
* Includes data from 784 currently depressed adults (vs. 727 controls) and 168 adults with remitted depression (vs. 178 controls).
* The study reveals moderate deficits in current depression across all domains tested (memory, attention, and executive function), with the exception of simple reaction time. Following remission, the memory deficits were small to moderate. They conclude that broad cognitive deficits are common in depression, with deficits in executive function and attention particularly persistent even in remission.
* A valuable point: this study found no relationship between cognitive deficits and depressive symptoms, and they note that even when such associations are found, they typically account for no more than about 10% of the variance in the cognitive disturbance. Thus, the cognitive side of depression is clearly not just about mood disturbance (or cognitive disturbance accessible to consciousness, such as rumination).

4. **Dalgleish et al. (2007)**

* A bit different, this one focuses on deficits in executive function as a cause of the autobiographical memory deficit in depression.