Thoughts:

0. Don’t respond to the reviewers in order, respond to IGD first since he asked for the most and the changes made in response to him are the most consequential.

1. We need to emphasize two effects: Cue x Task and Group x Cue x Task. Cue x Task because it is very strong, novel, and in line with NPS focus on basic work and advances in cog neuro.

2. A possibility for Fig 5: collapse across groups for the Q-S/animacy plot, since there is no Group x Cue interaction there. That will also bring out the nature of the Cue x Task effect. Probably worth computing Q-S/mobility across groups too (I bet we have both these graphs already and I’m just forgetting it).

3. For the question about distance between channels, here is the output of the spatial\_neighbors function with dist\_bw\_chans = 4:

max\_dist value of 4 corresponds to an approximate distance of 4.06 cm (assuming

a 56 cm great circle circumference head and that your electrode coordinates are based on an idealized spherical head with radius of 8.774327).

Min/Max distances between all pairs of channels (in chanlocs units): 1.460951/19.205354

Median (semi-IQR) distance between all pairs of channels (in chanlocs units): 12.082978 (3.162924)

Mean (SD) # of neighbors per channel: 6.2 (1.5)

Median (semi-IQR) # of neighbors per channel: 6.0 (1.0)

Min/max # of neighbors per channel: 1 to 9

Song et al. (2015) (Don Tucker’s group), in *J Neurosci Methods*, indicate that the mean distance b/w channels with the 128 channel EGI cap is ~2.7 cm. To test how sensitive our results are to this variable, I re-ran the MDD vs. HC Q/MI contrast with dbwc = 2 and got no significant results (sensible b/c that is lower than the mean distance and spatial\_neighbors indicates a min/max # of neighbors from 0-3), and with dbwc = 3, which gave me no significant results from 400-800 but a significant cluster from 800-1400 ms. Looking at the plots, the lack of significant effects from 400-800 ms is unfortunate because there is clearly a strong finding there, and follow-up tests on the waveforms in that time-window from electrodes identified by the dbwc = 4 analysis yield highly significant findings. Thus, it seems to me that dbwc = 4 is an appropriate setting for these data, keeping in mind that a mean interelectrode distance of 2.7 cm reflects data from areas where the spacing is much tighter than over the parietal scalp, where stretching is pronounced (e.g., over the occiput there is no stretching and the mean distance is notably smaller, which brings down the mean). (I’ve added a line to the text on this point.)

Visual inspection of Figure 6 should also limit concern about this, for two reasons. First, you can see that the clusters of significance closely follow the topographies where electrode-level significance is high—that is, you don’t see clusters that have a few electrodes in regions of high significance that spread to areas of low significance simply because the electrodes are proximal to each other. Second, you can see how strict the clustering assumption is by looking at the left parietal effect from 800-1400 ms and again from 1400-2000 ms; there is clearly strong activity there but it is very focal, and apparently not big enough to be considered robust. Thus, I think if one were to have a concern about this method it might be that it is overly conservative, not anti-conservative. But in this case we feel it’s worthwhile because the method provides the critical attribute—namely, a principled way to look across the whole head (important because we don’t know where the between-group effects may be).

4. Starns & Hicks (2005) include a “guess” response (so it’s not that unusual . . . I’m guessing I grabbed it from Simons and colleagues too so check there as well). “We also included a *guess* response because participants certainly should not have source information for items that they recognized only on the basis of guessing.”

5. I am looking for any other paper where a deep encoding condition is associated with worse performance in a conceptual vs. perceptual retrieval test, as is the case for us with respect to animacy words (worse for Question vs. Side). That’s an interesting an odd result, because you’d think that deep processing at encoding would support good conceptual retrieval—you had to think about the semantic properties of the object denoted by the word to answer the encoding question, but you didn’t have to think at all about Side and the Side placement is totally arbitrary, so you’d think you’d get Question > Side. Since you do not, I think this result must reflect interference at retrieval . . . you encoding the words okay in the animacy task, but when you respond to the Question condition you get mixed up vis-à-vis the mobility judgments and so that drives Question accuracy way down. Is there any precedent for that in the literature? Well, in Starns vs. Hicks (2005) Experiments 1A and 1B they did not control encoding strategy but they do show better source memory for font sizes than locations . . . but it’s not like font size is a semantic property of the words so I don’t think this really helps. This paper is a difficult read but its point is simple; in multi-source retrieval experiments, such as ours, people retrieve information about the two sources independently, and retrieving information about one source does not seem to cue memory for the other source. In their experiment 2, for instance, people learn two bits of source information for each word (font and location), and are either tested on both sources at one time, or at separate times (i.e., retrieve the font for every word, then go through again and retrieve the location for every word). You might have thought that you’d get some source-to-source cuing in the first condition, but you don’t—performance for either source is basically identical across these two testing formats. The bottom line is that they find a robust correlation b/w memory for the two sources—if you remember one accurately you’ll probably remember the other accurately—but they find no evidence that remembering one source helps you remember the other. It seems more likely that the correlation just reflects good encoding of the entire episode. **Actually, I think our data are consistent with this. Side information comes out just fine no matter what the encoding task, but Question information is affected . . . so they are separable bits of information about the episode.** “These results demonstrate independence in memory for different source dimensions in the sense that remembering one dimension does not alter the probability that other dimensions will also be retrieved.” “By demonstrating that an encoding variable increases memory performance on one source dimension without affecting another, the results of Light and Berger and of Marsh et al. suggest that contextual features are encoded and retrieved independently.”

Vogt and Broder (2007) provide a nice summary of Starns & Hicks (2005); when a source cue matches the encoding condition, source accuracy along that dimension improves—but critically, this benefit does not spread to the other source dimension. However, V&B argue that these results may be partially contaminated by the use of the average conditional source identification measure (ACSIM). Thus, they go on to use more complex procedures to address this issue, and the bottom line is they replicate S&H.

**Response to reviews of**

***The impact of depression on brain activity during source memory retrieval***

We were delighted to receive positive feedback from the reviewers of our work, who wrote that “Barrick and Dillon present an excellent study of source memory retrieval in major depressive disorder” (Reviewer 3), that “The manuscript is very well-written, the aims of the study are clear, and the analyses are thorough” (Reviewer 2), and that “This is an interesting topic and a novel design with potentially informative outcomes” (Reviewer 5). We also appreciate the reviewers’ constructive criticisms. Below we provide point-by-point replies to each one; we have taken the reviewers’ critiques very seriously and done our best to address them in this substantially revised manuscript. We believe the paper is significantly improved, and we hope the reviewers will agree.

Replies to Reviewer 1

1. *The selection and matching of the HC and MDD groups is a strength. Do the effects still hold when covarying education or IQ*?

Thank you for this question. As you noted, the groups are closely matched and do not differ on years of education or IQ, as estimated by WTAR scores.