Reviewer #1:

First, let me acknowledge the significant revisions the authors have made in this version of the manuscript. It remains an impressive study and the revision has certainly improved the coherence of the text and rationale of the research. In many ways, this revision is a leap forward, but the extent of the changes also means that other issues are becoming more apparent and some attention must now also be directed to more specific details to make sure the manuscript meets the standards of publication.

**Thank you!**

The general points raised in my previous review have been addressed, although the description of one issue remains ambiguous. Throughout the manuscript, the bias is interpreted as the participant's "over-estimation" of some form of relatedness. It could equally indicate the underestimation of relatedness in the measurement derived from word association. I would urge the authors to be more explicit about this in both the introduction and discussion.

**We clarified this point and added the measurement consideration in the introduction:**

**We use the term bias to indicate the overestimation of ratings for weak to moderately related pairs, as described in @Maki2007. However, the original @Maki2007 study used the @Nelson2004 norms as a metric to measure against, and measurement bias likely also exists. As mentioned earlier, these weaker associates may be underrepresented in the data with the one response free association task; thus, lowering their estimates and making participant estimates appear upwardly biased. By using the larger SWOW data, this study can explore if overestimation bias persists with less measurement bias by using the continuous response association set.**

**And in the discussion:**

**The SWOW norms size and construction lessens the measurement bias in the data, and these results support that some overestimation bias likely exists beyond potential measurement bias, especially in line with the traditional judgments of learning literature.**

The main issue with the current manuscript concerns the presentation of the findings. There are a couple of ways to improve this, and a detailed list is given below. I would especially like to draw attention to the abstract, as I found this to somewhat hard to digest. For example, the first sentence is difficult to understand without already knowing what's coming next, and the last concluding sentence in the abstract is a bit superficial.

The introduction is well-written for the most part, although the revised manuscript reduces the necessity to still include certain paragraphs that might be leftover from the first draft. First is the treatment of the connectionist models. It seems that this section on page 6 can easily be removed. Second is the discussion on WordNET measures, which also doesn't seem to be that consequential to the remainder of the paper. If you do feel these are important, then these sections will need to be integrated further in the introduction. I haven’t removed it yet, but I’m fine taking this section out if you are.

A final issue concerns the reporting of some of the results on page 19 and onwards. In these analyses, variability due to the nature of the JOR task (associative, thematic or semantic) is accounted for statistically, despite the fact that these tasks are also qualitatively interesting.

As a consequence it's difficult to understand how the interactions between direct and indirect relations vary across the three different JOR tasks. Unless there's a strong reason to do so, I would suggest splitting the analyses for each task separately as the extra statistical power does not outweigh the difficulty of interpreting findings from a mixture of qualitatively distinct tasks. This could for example also lead to a Figure 2 with three panels.

**This comment mostly affects Hypothesis 2 and 3 – the others are split by task already. We had envisioned these as examining the cognitive process, but the reviewer has a good point. We left in the original overall analysis, but then also paneled the analysis by task type (figure 2 and 3). The entire table of predictors for both of these analyses split by task are available on our OSF page.**

Detailed comments

page 4. "Direct word associations are 69 traditionally viewed as the context-based relation between concepts, usually found in text or 70 popular culture (Nelson et al., 2000)." I don't think this is the traditional view, at least not since the fifties (cf work by Deese).

**Fixed. Changed to say:**

**Direct word associations are traditionally viewed as the probability that the first word in the pair will cue the second.**

page 7. Please provide a reference regarding the better semantic relatedness relative to feature production norms

**Fixed – we thought the wording of the second sentence made this clear, but we added the reference to the end of this sentence as well.**

page 7, line 153 "Discussion of these measures then leads to the question of whether each one is truly assessing some unique concept or if they simply tap into various elements of our overall linguistic knowledge."

Why would judgments of relatedness only tap into linguistic knowledge? They could equally tap into non-linguistic knowledge.

**We clarified this section.**

page 8. The discussion of different effects based on forward and backward strength for the Judgments of Learning task suggests that both measures might be relevant to the current study. Why was only forward associative strength included? I suppose we could remove the section on backward strengths if we had to. We could also just add something in about only using only forward associations because we just wanted to have one measure of association that was directly comparable to the JAM ratings.

page 8. "backwards strength (i.e., the likelihood of the target when shown the cue)". Could it be the other way around: the likelihood of a cue given the target?

Fixed. We had this backwards

page 8. "However, the indirect relation between bird and feather is .063."

Does the indirect relation refer to cosine overlap? Please clarify.

**We changed this to say:**

**However, the semantic relatedness between the two is .063 when indexed using SWOW’s indirect association norms.**

page 9. I sometimes got confused about the judgment of associative memory (JAM) and the judgment of relatedness. Are these the same or do they differ (e.g. in terms of instructions). Please clarify this from the beginning.

**We added a sentence near the end of the introduction stating that the JOR task is a direct extension of the JAM task. (In the paragraph that starts ‘The present study expand upon previous JAM studies…). We added this sentence to the end of that paragraph:**

**Thus, the ensuing JOR task is a direct extension of Maki's (2007a) JAM task.**

**Additionally, we added this sentence to the next paragraph that starts hypothesis 1:**

**Because the judgment task we employ is an extended JAM task that also includes semantic and thematic judgments, we subsequently refer to this task as a judgment of relatedness task (JOR), regardless of which judgment type is being elicited.**

page 9. "Overall, the JAM function has been shown to be highly reliable and generalizes well across multiple variations of the study, with item characteristics such as word frequency, cue set size (QSS), and semantic similarity all having a minimal influence on it (Maki, 2007b)."

I'm a bit confused by the last part of the sentence: why would the bias not be impacted by the semantic similarity of the items? If they are more similar, shouldn't the bias be lower?

**We reworded this a bit:**

**Overall, the JAM function has been shown to be highly reliable and generalizes well across multiple variations of the study, with item characteristics such as word frequency, cue set size (QSS), and semantic similarity all having a minimal influence on the function [i.e., similar intercepts and slopes were found for manipulations of these variables, including semantic similarity of the word pairs; @Maki2007].**

A screenshot of text

Description automatically generated

**In the original study, when semantic similarity of the pairs was manipulated with JCN, the same JAM function was found for low and high pairs – thus, semantic similarity was not thought to influence the slope. This analysis suffers from being a simple dichotomy rather than a predictor like we have done, but these are the results we are referencing.**

page 10. "Further, the influence of 219 indirect relations and their potential interaction on judgments have not been investigated."

This is not entirely correct, as this was one of the main research questions in De Deyne et al, 2013.

**We clarified this to say:**

**Further, the influence of indirect relations on judgments has not been investigated within the context of multiple judgment types (but see De Deyne et al., 2013 for a review of both SWOW association types within the context of semantic similarity judgments).**

page 11. "Next, we explored the predictions from semantic network models that the relation

between associations and semantics would be bidirectional in nature (i.e., both types of

knowledge interconnected in memory)."

I'm a bit confused by this sentence, How can associations and semantics be bidirectional? It seems that semantics (or at least how words are related) can be thought of as the combined effect of one (direct) or many (indirect) associative links between word pairs, similar to distributional models.

Can you please rephrase this?

**We changed it to say:**

**Next, we explored the predictions from semantic network models that the relation between associations and semantics would be interconnected by nature (i.e., both types of knowledge closely linked in memory).**

Along the same lines, I would rephrase the following sentence as "differently interactive" is a bit vague: "Because judgment and recall are different cognitive processes, we used this hypothesis to examine how memory networks may be differently interactive for memory in comparison to judgment."

**We rephrased to say:**

**Because judgment and recall are different cognitive processes, we used this hypothesis to examine how the interactive structure of memory networks may differ based on task.**

page 13. In Table 1 the values based on the Florida norms are included, whereas later in the article SWOW values are used. I understand that the former data were used for selecting the stimuli, but it might be useful to provide the descriptives based on SWOW as well in this table.

**These values were included in Table 1 - they are labeled direct association and indirect association.**

page 16. It would be interesting to know the correlations between the three JOR version tasks. Does the sentence at the end of the paragraph ("r < .26") refer to these? If not, can you add them?

**We rephrased this section to clarify what correlation(s) this indicated**:

**The correlations between judged value, recall, direct and indirect association were all \*r\*s < .26. These correlations were similar regardless of judgment condition.**

**The suggestion here is unfortunately unclear – if the request is judgment and recall scores, these are low rs ~ .07 and approximately the same between tasks. If it was somehow correlating performance on condition, we were unsure what would be correlated between conditions, as words are different for each person in each task.**

page 18. "Overall, 407 thematic JORs had the highest intercept at .61, while JORs elicited in the semantic 408 condition had the lowest intercept at .52."

This seems inaccurate, as both the associative and semantic condition had an intercept of .52 (see Table 2.

**We updated the text accordingly, as semantic and associative were equal but lower than thematic.**

page 18. "Last, the semantic judgment condition showed that both direct and

415 indirect relations were important (as this judgment type had the highest indirect

416 contribution of the three conditions)". Can you add the numbers to support this statement in the text?

**These numbers were added to the text.**

page 19. "The goal of the next analysis was to test for an interaction between direct and indirect

430 association when predicting participant JORs." I would unpack this and reiterate the research question (what are you trying to investigate) and the hypothesis to be tested that goes with it.

**We reworked this to say:**

**Next, we sought to test the interactive relationship between associative and semantic overlap. If this interactive relationship exists, a statistical interaction should be detected between the database norms when predicting performance on the judgment task. As such, the goal of next analysis was to test for this interaction between direct and indirect association when predicting participant JORs.**

page 20. Can you explain what the role and interpretation is of the intercepts for associative, semantic and thematic intercepts? Why does higher bias result in higher recall? In particular, I was wondering if you have an explanation why the associative intercept was insignificant whereas the semantic and thematic ones were.

**We updated this section and added some thoughts here:**

**For direct association, increasing sensitivity to the relation between pairs lead to increasing likelihood of memory, which is not surprising. Indirect association also showed this effect, that stronger indirect sensitivity to word pair relation also increased memory in the thematic and semantic judgment conditions, similar to indirect memory results from @Huff2011 and @Huff2012b. The intercepts or bias estimates from the first hypothesis indicated that increasing participant overestimation of weakly related pairs also predicted increased recall. Potentially, this result can be viewed as self-fulfilling, the more related participants thought the weakly related word pairs were, they more likely they were to remember them.**

page 21 "In contrast to De Deyne et al. (2013a), we found bias in judgments for pairs with no direct relation across all three judgment conditions (average judgment = 50.36); however, only 5 pairs were available for comparison to their results."

I can't follow here. Do you mean that only 5 pairs did not have an indirect relation in the current study? Also, one still expects bias to be present. As mentioned in the previous review, the basic idea is that some of this bias ("over-estimated JOR" ) reflects a compression effect as forward strength might only capture strong pairs and discard weak pairs. If relatedness or indirect association is taken into account the intercept should be lower (but perhaps not absent).

**Correct – the study was designed to capture pairs with low-med-high relations, so our study plan did not include zero overlap pairs. When updating to the SWOW norms, we had 5 pairs with zero overlap. We included this section to help address this question from the previous reviews, and updated it a bit to help clarify:**

**Finally, in contrast to the study conducted by @DeDeyne2013a, we found bias in judgments for pairs with no direct relation across each of the three judgment conditions (﻿average judgment = 50.36); however, these findings should be viewed cautiously as our stimuli contained 5 item pairs that had no direct association. The SWOW norms size and construction lessens the measurement bias in the data, and these results support that some overestimation bias likely exists beyond potential measurement bias, especially in line with the traditional judgments of learning literature.**

page 22. The other reviewer also alluded to this, but perhaps instead of referring to numbered hypotheses in the discussion, it would help the reader's understanding if you refer to the research question instead.

**In each reference to a hypothesis in the discussion, we included the hypothesis or test description with it. In the first round of edits, we restructured these to keep the numbering to flow from methods 🡪 results 🡪 discussion, while also making sure to reference their goals with each sentence.**

page 32 - 34. Please improve the structure of the table so main effects and interactions are clearly indicated. Consider improving readable by using the same naming conventions used throughout the manuscript to indicate the variables (e.g. direct strength, indirect strengths, etc.). Finally, provide a bit more information in the table caption about what research question the table is addressing and what it shows. For Table 5, please rephrase the Note and spell out the JOR judgment types in full in the table.

**We have edited these tables accordingly.**

Reviewer #2:

I found this reworked revision to be much improved. The reviewers' comments were carefully considered and addressed in the revision. However, I still see the need for further elaboration in the Discussion section. Specifically, more discussion is needed for why direct association but not indirect association was such a strong predictor in the associative condition.

**We added a few sentences to the discussion on this point:**

**The observation that direct association was the strongest predictor of both judgments and recall within the associative condition and that the indirect association was strongest for the semantic and thematic conditions is not surprising. Direct associations are designed to capture the associative overlap shared between word pairs whereas indirect associations are thought to tap into elements of the overall semantic network and represent similarities in meaning rather than cue-target probabilities. Therefore, these results appear to reflect the task demands for each judgment condition.**

To address this point, the authors could relate their findings back to those in the mediated priming literature that showed semantic priming (e.g., Lerner, Bentin, & Shriki, 2012; Jones, 2010, 2012) and memory (e.g., Huff & Hutchison, 2011, which the authors have already cited; Huff, Coane, Hutchison, Grasser, & Blais, 2012) for indirect associates only under limited conditions or tasks. Likewise, why was indirect association a stronger predictor for the semantic and thematic conditions in comparison to the associative condition?

**We added information on how our results correspond to the findings you mention here:**

**This finding may also be comparable to results in the semantic priming literature, wherein direct and indirectly related pairs show different priming effects [@Lerner2012], often modulated by task [@Jones2010; @Jones2012a], and recognition is also too influenced by indirect relations [@Huff2011; @Huff2012b].**

On a more minor note, there were several typos throughout the paper that the authors need to fix. Overall though, I think this research is interesting and relevant to both semantic priming and associative memory researchers.

**We reread the paper for typos and updated accordingly.**