**Reviewer 1**

This paper explores how different types of information contribute to processing meaning. The paper extends previous work in a number of interesting ways. First, the authors explicitly relate judgements of relatedness with recall performance and second, they investigate the nature of the the semantic relation between two words in a pair and how different semantic variables interact. The analyses are carefully conducted, and the findings, methods, and procedures are mostly well explained. Making the data and scripts available on OSF is also greatly appreciated.

The results themselves point to a privileged role of associative and thematic information in semantic processing, over semantic information captured by property overlap. Although the implications of the current results are of potential significance to the field and the empirical work and analysis are of high quality, I have some reservations about the framing of the research question. More precisely, it seems that this study (and admittedly a number of studies before it) are comparing apples with oranges by looking at language-based and psycho-experimental estimates (word associations, property listings) while at the same time also using different metrics to determine the similarity between them (e.g. direct strength or distributional overlap).

A central premise of this work is that there are indeed three different ways in which words can be related, either because they co-occur in language (what we could call first-order co-occurrence or syntagmatic relations), or they are related words that occur in similar contexts (second-order co-occurrence or paradigmatic relations), and then finally some other form of relatedness not captured by language itself, but captured by semantic properties (a variant on the word association task).

In various parts of the paper the authors rightly indicate that these three components of meanings are by no means independent. However, in other parts, the nature of these relations is presented in a way that somewhat contradicts this. For example, on page 4, they write that word associations typically arise through their co-occurrence in language. This is surprising as word correlations mainly capture semantic properties not consistent with lexical co-occurrence. For example, they also include a great number of paradigmatic relations (e.g., dog - cat). Again, this is a common misunderstanding which has been echoed in other studies where rather weak correlations between lexical co-occurrence and word association has been taken as strong evidence that word associations indeed primarily reflect lexical co-occurrence. A similar problem occurs in specifying meaning in text-based representations. For example, LSA is introduced as a measure of thematic relatedness. This might be

merely an issue of wording, but the relatedness from LSA is not necessarily thematic, it merely captures distributional overlap (second-order representations) based of the context in which words co-occur.

The second point is that the idea of studying associative strength in isolation might not be a fruitful enterprise. James Deese made this case about 50 years ago (Deese,1965), arguing that the interesting aspect of word associations are captured by the distributions or larger structures in which word associations are embedded instead of looking at strengths between two words in isolation. In other words, it's difficult to reconcile looking at direct forward or backward strength without considering indirect paths connecting two words. This is also what LSA and other techniques aim to do: they track second-order occurrence through indirect co-occurrence through SVD to capture what in the manuscript is called "thematic" relatedness (although here as well, LSA will contain a mixture of both first-order and second-order relations, so this term is not entirely appropriate). The bottom line is that this suggests that association and relatedness are at opposite ends of a continuum rather than being categorically distinct.

At this point, I believe there are two sensible strategies. One would be to investigate first-order association and higher-order associations by either comparing direct association from word association norms and derive indirect associations by comparing distributions between associative norms (for example using cosine overlap). An alternative or separate approach would be to use lexical co-occurrence strength (e.g. point-wise mutual information of words occurring in a sentence) from text corpora and compare it with distributional relatedness. This would allow you to disentangle two aspects of meaning, and capture this continuum spanning association and relatedness without having to assume a three tier system as proposed in the paper.

Beyond reframing the semantic measures, a more in-depth discussion of the meaning of the JAM function would also strengthen this paper. For one, it seems that the intercept in the JAM function represents a compression effect instead of a subject over-estimation bias (see De Deyne & Navarro, 2013) where we might assume that word associations are biased because they don't capture weak strengths due to the small number of responses in the Nelson database. In that case, we might find that including more associations or including indirect paths might be able to account for variability in the participants relatedness judgements for items that now have zero forward strength. This is at odds with interpreting the JAM function as overconfidence in judgement for the intercept (see p 22).

Furthermore, the paper could do a bit more to unpack the theoretical framework and implications of the interaction effects discussed in this work or indicate where the work is purely exploratory. This will require to carefully assess speculative aspects of the work, and flag these as such. In particular, the possibility of a three-tier system of semantics seems highly speculative once the interrelatedness of the measures is fully appreciated. Along the same lines, the discussion about time-course of semantic access (see p 24) also seems somewhat at odds with proposals for early linguistic access and late conceptual processing (cf work by Larry Barsalou on LASS theory, and recent work by Max Louwerse).

Finally, one of the real neat aspects of the paper, namely linking the recall to the judgements of relatedness could be explored more extensively, but here as well, the findings might well depend on whether the intercept of the JAM functions indicate overconfidence or a compression effect due to a small number of (direct) associations being considered) and whether word association data or text-data is used to investigate association vs relatedness.

**Reviewer 2**

SUMMARY AND EVALUATION: The authors investigate the interactive relationship between semantic, thematic, and associative word pair strength in the prediction of judgments of learning (more specifically, judgments of recall, JOR) and cued-recall performance. Based on their results, the authors propose a three-tier triangle model that explains how thematic and associative information is incorporated with featural similarity. The literature review in the Introduction was thorough and the study was well-designed. However, the manuscript needs substantial revision, particularly in the Discussion section. Below I list points related to this concern in more detail.

1. The motivation of the study was not as clear as it could have been. As a reader of any research article, I want to know early (within the first two manuscript pages or within the first half page of the published paper), why the study is important and what the purpose of it is. What gaps in the literature does it fulfill beyond other studies that have examined the interplay of all three (or just two) of these variables? Why is it important to evaluate the interaction of semantic, associative, and thematic information in judgments of recall and recall performance? I did not get these answers from the first couple pages. Rather the authors present a brief history of paired-associate learning and a description of word-norming. It wasn't until page 7 that the purpose of the study was explicitly stated.

2. Lists of hypotheses are often included in theses and dissertations though not typically in most published papers. For greater coherence, include them each as a paragraph, relate them back to the purpose, and label them by the main point of the prediction or research question (e.g., Prediction of Judgments of Relatedness rather than "Hypothesis 1." I tried fixing this. I gave each hypothesis as title in the results section and turned the descriptions of them into paragraphs rather than a list in the intro.

3. I confess that I am not too familiar with the choice of analyses used by the researchers. However, I would like to see some more statistical support (ANOVA results) on p. 16 for the claim that the mean JOR for the associative condition was lower than the sematic and thematic conditions. That certainly appears to be the case from the descriptive statistics provided, but adding the ANOVA would nicely verify that. Also, add ANOVA for the recall results to show the lack of difference across the three conditions. Done!

4. The Discussion section was a little weak in that further elaboration was needed. For example, on pp. 22-23 "Additionally, this finding may be taken as further evidence of a separation between associative information and semantic information, in which associative information is always processed, while sematic information may be suppressed due to task demands (Buchanan, 2010; Hutchison & Bosco, 2007). Elaborate on this point futher particularly what is meant by "task demands." Hutchison and Bosco used a letter search task. How does your task also suppress semantic information?

5. Given that the "study did not explore the timing of information input from each of these systems" (p. 24) should the proposed model be described in steps (e.g., First, Next, Finally)? The first step seems to contradict the possibility that semantic information was suppressed in the JOR and cued-recall tasks.

6**.** Ending a paper with "further studies will be needed to fully understand…" is lackluster. Instead,state more specifically what the next steps (studies) could be in this line of investigation then maybe reiterate the overall contribution of your study to the line of research. Edited the final paragraph

**EDITS TO MAKE**

**Reviewer 1**

1. Clear up contradictions with meaning being independent or not independent
   1. “In various parts of the paper the authors rightly indicate that these three components of meanings are by no means independent. However, in other parts, the nature of these relations is presented in a way that somewhat contradicts this. For example, on page 4, they write that word associations typically arise through their co-occurrence in language. This is surprising as word correlations mainly capture semantic properties not consistent with lexical co-occurrence.” Tried addressing this. Mostly just attempted to reiterate that each component is indeed related and that co-occurrence is just one way in which associations arise.
2. “At this point, I believe there are two sensible strategies. One would be to investigate first-order association and higher-order associations by either comparing direct association from word association norms and derive indirect associations by comparing distributions between associative norms (for example using cosine overlap). An alternative or separate approach would be to use lexical co-occurrence strength (e.g. point-wise mutual information of words occurring in a sentence) from text corpora and compare it with distributional relatedness. This would allow you to disentangle two aspects of meaning, and capture this continuum spanning association and relatedness without having to assume a three tier system as proposed in the paper.” Not sure the best way to approach this
3. Discuss more about the meaning of the JAM function (see the Maki papers) Added a paragraph on JAM ratings and free association norms not being calibrated, even though they should in theory be measuring the same thing.
   1. Compression effect vs overestimation (De Deyne & Navarro, 2013)
4. LASS theory (barsalou, louwerse) (in regards to timing of information on p. 24) . “Along the same lines, the discussion about time-course of semantic access (see p 24) also seems somewhat at odds with proposals for early linguistic access and late conceptual processing (cf work by Larry Barsalou on LASS theory, and recent work by Max Louwerse).” Added a paragraph on LASS theory. Might need more.

**Reviewer 2**

1. Make the motivation for this study more clear
2. Change how we introduce the hypotheses
   1. They suggest that each should get a paragraph that is labeled by the main point Tried addressing this.
3. ANOVA comparing statistical differences in JOR scores are across judgment conditions and an ANOVA showing a lack of difference in recall across pair type Done!
   1. Add this in to the results Done!
4. Update the discussion with more elaboration. Expand on “task demands” (p. 22-23)
5. Elaborate on timing of information as mentioned on p.24 This was hopefully addressed with the LASS theory stuff from reviewer 1
6. Expand the final paragraph/conclusion Tried updating the conclusion