Submitted Paper Revisions Memo

To: Dr. Melissa Jungers, Associate Editor

From: Brendon Mizener

CC: Drs. Mathilde Vandenberghe-Descamps, Hervé Abdi, and Sylvie Chollet

Dear Dr. Jungers and Reviewers:

We are deeply grateful to all of you for your very insightful comments on our draft. These suggestions have been instrumental in helping us find perspective on our manuscript and incorporate a new sense of clarity. We feel that your suggested revisions have helped us clarify aspects of our paper that were unclear or confusing and that our paper is now much stronger.

You noted in your decision letter that there were three general points that concerned the reviewers. The first point concerned the analyses. We did not reduce the number of analyses, but we have streamlined the text of the document so that the figures and text were not redundant. We have also included a new section (“Why these methods?”) in the general discussion that offers further justification for some of the less familiar analyses, as well as a new table (Table 3) that shows the analyses we included, some similar methods, and possible applications.

The second point was well made—the word “cognitive” was not the best term to use given the overall context. We have edited the paper so that “cognitive” is replaced throughout with “qualia.” We feel this more accurately reflects the concept behind the analyses.

The third point identified a concern about participants. We have added a table (Table 2) that includes participant demographic information. However, Reviewer 2 makes the specific point of asking which languages were spoken by the participants. Unfortunately, this was not part of the original data gathered as part of this study and we regret that it cannot be reported. The other point, about psycholinguistic literature, has been addressed in the text.

Below are our changes or rebuttals to each of the points made by the reviewers.

Reviewer 1

General Concerns

There is a lack of statistical results. I do understand that the authors explained this in the introduction with their rationale for using dimension reduction procedures, but these types of procedures work best when paired with some type of inferential statistic (frequentist or Bayesian) to help interpret the results. I make liberal use of multidimensional scaling (MDS) in my own work because it is both useful and provides far more information than it often is credited for. However, the end result cannot be the resulting cognitive space, even with confidence intervals imposed around the centerpoint. The standard errors can be fed back into a formula to obtain observed values that can be compared to critical values which, as this study is really a two by two design, be corrected for multiple comparisons. There are also some new methods to use the cognitive space derived from MDS or other similar procedures in hypothesis testing (see Patten & McBeath, 2020 and Patten, McBeath, & Baxter, 2018).

* The reviewer makes and important point. We say in the introduction that multivariate analyses *in general* can reveal cognitive spaces, however, we did not mean to imply in our discussion of the MDS specifically that the result of the MDS is a cognitive space. What we intended to say was the MDS shows the distances between participants, and that finding and bootstrapping group means helps us to evaluate possible group differences. We hope this has been clarified appropriately in the text.
* In response to this comment, we also performed a *t*-test of the first dimension factor scores for the MDS to identify differences between participant groups, the result of which supports our conclusion about group differences.

There seems to be too much information at times. Certain elements of the data transformations, such as constructing the brick, are unnecessary and might confuse some readers. Likewise, the same is true for figures elucidating these procedures.

* We appreciate this comment, but we felt it was important to keep the description of this process intact so that readers who are interested in pursuing these same analyses have a guide to the process. We feel that the procedural aspect of this paper is vital because the paper is focused on the process and uses the data and the results of the analyses more as a case study than as a result in and of itself (although the results are both interesting and significant). We have, however, clarified and streamlined the results sections of all three Experiments to reduce the amount of information and make it overall more comprehensible.

There may be too many analyses. It would be a more understandable and comprehensive read if only some of the multivariate approaches were used.

* We appreciate this comment and apologize for writing the initial draft in a way that was less than coherent. We hope that with this updated draft we have clarified and justified our reasoning for the methods we used.

Correspondence analysis, as far as I'm aware, assesses only two dimensions. It's possible, if not likely, that the data in this work exist in several dimensions. The MDS procedure is not limited to two dimensions and will provide a Scree plot similar to the one used in the paper that will allow the authors to pinpoint the number of dimensions necessary for this data. Using two dimensions for a data set that exists in three will produce erroneous correlations and proximities. Likewise, reporting both hierarchical cluster analysis and MDS can be condensed by using an additive tree cluster or embedding the hierarchical analysis into the MDS results a la Shepard (1980).

* This specific point about the differences between CA and MDS was helpful in guiding our revisions as to what people may want to know about CA specifically and about other methods generally. For example, we have clarified in the draft what the dimensionality of the CA is (the lesser of *I* – 1 or *J* – 1, where *I* is the number of rows and *J* is the number of columns). Another difference between the two methods that did not fit specifically in the paper, but may help clarify our point, is as follows: Whereas MDS only is only concerned with distance between or among observations—as in the way we used it for our analysis of participant group differences—CA computes the relationships between observations, between variables, and between both observations and variables.

Specific Concerns

There is some sort of issue with spacing throughout the manuscript. Page 2, line 7; page 3, line 24, line 27, line 32, line 46, line 53 and throughout. Usually it is a missing space, though there are instances of an extra space.

* This was an artifact of conversion between pdf and word that we thought we had addressed in the initial draft. Thank you for pointing this out, it has been fixed.

Page 3, line 7: No capitalization for "world."

* Thank you for noting this capitalization. The point we’re making here is that the events of the world over the past year (going on two years now) have been exceptional and warrant proper noun treatment. We have thus opted to capitalize “Events” as well.

Page 3, line 27: Remove "In 2004". Also consider the advice of Kail (2015), which urges authors to focus on the findings of studies rather than the people who found them. After all, the topic of the sentence should be the phenomenon in question and not the scientist. This sentence, then, might read, "Something as simple as the sound of a crunch when eating a potato chip influences the taste (Zampini & Spence, 2004)."

* Thank you for this helpful comment. I was unfamiliar with the work of Kail, but this has been a helpful resource. We have incorporated this change into the manuscript.

Page 4, line 7: This paragraph (or the discussion of MDS later on) would be more complete with references to Shepard (specifically 1962 and 1980, though others are applicable) and Kruskal (specifically Kruskal & Wish, 1978). Hout et al. (2013) may also be useful.

* Thank you for these suggestions – they have been added to this section and the MDS section, depending on where they are most relevant.

Page 5, line 24: Thompson's first name is in the citation.

* Thompson’s first name removed.

Page 6, line 38: Heading is inconsistent with others.

* This heading is intended to be a subheading, one level below the previous heading. If the reviewers/editors feel it is more effective as a heading one level higher, we’re happy to change for the final version.

Page 8, line 53: MDS can also analyze correlations and confusion matrices, in addition to distance. "Similarity" might be a better term.

* Changed to “similarity.”

Page 16, line 43: The RotCorr procedure (in Patten, McBeath, & Baxter, 2018 and available at [URL removed for ) can quantify the correlation between arousal and dimension 1. Feel free to reach out with questions on this if you decide to use it and have any.

* Thank you very much for your kind offer. Unfortunately, in this case, the excerpts were not rated on arousal, rather, “arousal” is an interpretation of the dimension that was extracted by the CA. In order to evaluate arousal in an unbiased way, we would need to run another survey, which, given the time constraints, was not possible.

Page 22, line 31: It would be nice to see values for the means and variances.

* Values for means and standard deviations have been added to the MDS results section.

Page 23, line 19: The relative closeness of factor scores is a good place to add statistical validity. Using each individual's point as a datum, a t-test can be calculated to give meaning to the distance between centerpoints.

* The partial factor scores obtained by the MFA are not group means, and thus cannot be used in hypothesis testing. They are individual factor scores extracted from the GSVD of the co-occurrence table for each group, projected into the space computed from the normalized table of both groups combined (concatenated). Additionally, unlike many other multivariate analyses, MFA should only be used for exploratory purposes, or as in the case we have here, where we have made inferences from the MDS and use MFA to evaluate where the differences identified by the MDS may be coming from.

Page 36, figure 2: This Scree plot is odd to me. I'm used to looking for a point of diminishing returns (i.e., where the curve flattens out) to determine the appropriate dimensionality of my results. Two dimensions looks like a bad fit on this plot. Three looks much better, and a case could be made for 6. I see that explained variance decreases with increasing dimensionality, but variance is not always correlated with meaningful dimensions (though it is a central assumption of PCA and other procedures) and stress between points (and, thus, the estimation of distances) should improve with increasing dimensionality.

* We regret our initial lack of clarity surrounding this scree plot. Because the dimensions in CA are orthogonal, visualizing dimensions beyond the first two requires more plots, and thus choices about what to display are necessarily conservative. We initially intended to include only the first two dimensions, for the sake of efficiency. We have since updated the scree plot to include significant dimensions as determined by permutation testing. We have also clarified in the text how the dimensionality is limited by interpretability in the case of CA. Also, please refer to our response to the point above about how the dimensionality chosen for CA and that chosen for MDS are different in a pedagogical sense.

Overall, I like this paper. I think the message of ecologically valid data being available in unusual circumstances is an important one. Likewise, the message of using multivariate procedures to help us understand confusing data is necessary and something I very much agree with. I think the results are interesting, as well. Unfortunately, in its current state, the work feels unfinished and unpolished. I think the sheer volume of analyses thrown at the reader needs to be restrained or combined into more easy-to-consume plots (for instance, instead of displaying all the excerpts that predicted a dimension, what acoustic features can be pulled from those excerpts that explain the dimension?). I'm recommending the journal not accept the work in its current state, but I urge the authors to resubmit after tightening the study up a bit.

* Thank you for your endorsement of the message of this paper. We hope that the revisions that we have made have sufficiently tidied and clarified the paper and that you find our revised version recommendable.

Reviewer 2

Based on my review of the manuscript, I found the content easy to read and follow. Overall, I found that the implementation of these techniques to be quite beneficial to, as the authors stated "add to the methodological toolbox." While I agree, conceptually, that these methods will enrich one's methodological toolbox (not just psychologists, these methods would benefit a number of social scientists), I found the arguments for when one would use these methods to be lacking.

* Thank you for your kind words. We have added a new section in the general discussion, (“Why these methods?”) that we believe clarifies the “why and when” argument about using these methods.

Moreover, implicitly and even explicitly, it makes sense that these types of methods are useful. However, these types of cluster analyses are common in the cognitive sciences (e.g., among psycholinguists) — e.g., like LSA (latent semantic analysis) — and most cognitive scientists are not stranger to these types of techniques.

* This is an excellent point. We have added a new table (Table 3) that lists the analyses we used in this paper, some similar analyses, and situations in which they are useful. We hope this helps to offer some context for what types of analysis are useful in this situation.

However, the techniques described and used in the manuscript do seem somewhat novel for the field, and it seems the authors missed an important opportunity to contrast how these methods are different for other common cluster analyses used in the cognitive sciences.

The manuscript would also benefit from a deeper discussion of when and why these methods would be used in perceptual frameworks, and how they may be useful for auditory research more specifically. I am not sure many perceptual psychologists who study auditory domains would be convinced by the arguments made - mostly because they need more theoretical backing. I just think a stronger justification is warranted and would only strengthen the manuscript.

* We agree with these two points and they gave us guidance on how to justify and provide context for the analyses we used. Please see our response above for how we have addressed this.

Another point, I am not sure that "cognitive music listening space" is really describing what it is that the authors are doing. Are the authors really describing cognition or are they describing an aspect of cognition. The authors even discuss how semantic space in one's cognitive system may be shaping the words and interpretations the participants were using to categorize the musical stimuli. I think the authors should consider the implications of describing their task as cognitive, when I am not sure it truly is. Or they need to make a very clear argument for why "cognitive" is better than something else.

* This is an excellent point. We have edited throughout the paper to replace “cognitive” with “qualia”, which we feel more accurately characterizes the responses that the participants made to the stimuli.

Finally, the authors should consider pulling from the psycholinguistic literature more to discuss the cultural differences in their semantic networks to better describe the differences between French and American speakers. It isn't enough to say, that it is likely the case, rather find citations to back it up. This will make the results more compelling and may even provide a context to bridge the gap between other domains, making the manuscript beneficial to other readers.

* Thank you for this constructive comment. We have added citations to Osgood and the Open Multilingual Wordnet which we believe offer support for the differences in adjective use between languages.

Other than that, I thought the manuscript was well written and easy to follow. There are some minor changes that need to be made, which include the following:

1. The authors describe the figures in the main body of the text in a way that makes the text and figures redundant. The authors should either remove the figures or they should provide a general description of the clusters, instead of providing an overly detailed description.

* Thank you for this comment! This comment helped us to streamline the results sections of all three experiments. We have removed much of the description, so that the resulting text is much more concise and focuses on the “big picture” so that readers can focus on the figures for the detail, which we hope is more intuitive.

2. There are many places in the document that words should be separated by a space or there are redundancies in words used. The authors should pay close attention to these typos and fix them.

* This was an artifact of conversion between pdf and word that we thought we had addressed in the initial draft. Thank you for pointing this out, it has been fixed.

Reviewer 3

Major concerns:

1. Participants. Demographic information should be included about the participants, such as average age, gender, years of musical experience, language(s) spoken, etc. Minimally, this information could be represented in a chart in the supplemental material section. The recruiting methods, with an emphasis on UT undergraduates in Experiment 2, could lead to participants who differ in more ways than musical expertise across the two experiments.

* Thank you for pointing this out. It was not our intention to be overly discreet with our participant demographics. We have added a table (Table 2) that includes much of this information. We also ran a separate analysis that included all participants who took the English language survey, including the participants who were originally excluded as a third group, that did not show any significant differences between the groups of English speakers. This plot (MDS) is included in the supplementary materials.

The participant loss should be addressed. In Exp. 1, only 27 of the 84 responses were included. In Exp. 2, only 278 of the original 520 were included after removing incomplete surveys and individuals who reported a nationality other than American (or an American-other nationality compound). The choice to include only complete surveys is a valid one, but the large number of participants who were not included should be mentioned in the discussion. What does this data loss tell us about on-line data collection?

- This is another excellent point. We apologize for our initial lack of clarity on how attrition affected our study. We intended to include this in the initial draft and we recognize that we were not as clear as we could have been. We have added more discussion of this specific issue in the general discussion under the limitations section. We hope that this addition is satisfactory.

1. The tasks in Exp. 1 and Exp. 2 are different and are performed by different populations. Please give stronger justification for directly comparing these results in Exp. 3. Is there literature supporting this type of combination? How might experienced musicians perform on Exp. 2? Would they be expected to use the adjectives in a similar way?

* We apologize again for not being as clear as possible with the justification for Experiment 3. We included a new subsection under experiment 3 (“Justification”) that we believe addresses these specific concerns. We have included a reference to Bigand and Poulin-Charronnat (2006), which suggests that experienced musicians and nonmusicians would not be expected to differ on the task in Experiment 2. Also, we appreciate you calling to our attention that we were unclear about the lack of group differences by musical training in experiment 2, which we mention in the results section of experiment two. We have also added a comment about that in the general discussion.

1. The methods of analysis are well described, beginning on p. 7. Although each analysis contributes a specific aspect or view of the data, it is not clear why so many methods are needed. Do some analyses offer a better picture, while other analyses contribute less to answering the initial research questions? It is worth emphasizing the contribution of each and why each is needed.

* Thank you for this comment! This helped us guide our writing in the new section under the general discussion (“Why these methods?”) which we feel addresses these specific questions.

Minor concerns:

1. It appears that only the American undergraduate students were compensated by being given course credit. Were the other participants compensated? If so, how?

- We have clarified this specific point in the methods sections of experiments 1 and 2. No participants besides the UTD undergraduates were compensated in any way.

2. Please check the spacing in the manuscript.

* This was an artifact of conversion between pdf and word that we thought we had addressed in the initial draft. Thank you for pointing this out, it has been fixed.