University of Oslo
Department of Linguistics and Scandinavian Studies
Dr. Timo Roettger

Dr. Daniel Simmons
Advances in Methods and Practices in Psychological Sciences

Date: May 13, 2021

Response to concerns about initial submission: AMPPS-21-0020 - Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human speech analyses

Dear Dan,

We would like to express our gratitude for your thorough first reading of our Registered Report. You point out two important concerns that we would like to address.

Concerns about the nature of the data set

You articulate the concern that the data set might be restricting our pool of possible analysts due to the fact that the data are from an experiment on German participants. Thank you for bringing this to our attention. We actually think that using a data set that is *not* based on English speakers reduces the average familiarity of analysts with the specific phenomena under investigation and thus allows for a less biased assessment of the data. In other words, we see this as an advantage rather than a disadvantage. The language sciences are still very anglocentric, leading to a high familiarity with patterns in English. A dataset on English speakers might amplify analytical decisions based on prior beliefs. The question of familiarity with cultural or linguistic properties of a dataset is in itself an interesting research question but outside the scope of our present study. From a technical point of view, speech scientists are trained to analyse acoustic data of unfamiliar languages, so we do not anticipate people being hesitant due to a German data set. Moreover, we believe that analysts who are "blind" to the meaning of the recordings are arguably less biased. In order guide analysts, our metadata provide detailed descriptions of the relevant components of the data set. We have added this argument to the Registered Report. We do acknowledge that familiarity with German is a

E-mail: timo.roettger@iln.uio.no
www.uio.no

variable of possible interest, thus we also added a couple of items to our surveys that assess this information: We are now asking analysts to specify their proficiency in German and whether they have worked on German speech before.

Concerns about the sample size

You further mention concerns about the sample size referring back to the seminal paper by Silberzahn et al. (2017). We believe, despite the exploratory nature of our study, that these concerns are relevant to whether our meta-analysis will actually be able to quantify variability between teams to an appreciable degree of certainty in the first place. As outlined in the manuscript, our primary aim is to assess the variability of the reported effects, rather than the meta-analytic estimate of the investigated effect *per se*. To estimate the degree of uncertainty around effect variability as driven by number of teams, we ran a series of sample size simulations with values of variability extracted from Silberzahn et al. (2018). Variability among teams was operationalised as the standard deviation of the teams' reported effects from Silberzahn et al. (2018) (which we *z*-scored prior to simulations to make it comparable to our study). For the mean of the teams' true standard deviation (0.68 *z*-score), the simulation indicates that the degree of uncertainty around the estimated teams' standard deviation will be below 1 SD at any sample size greater than 10 teams. The code for the sample size simulation can be found here. (Incidentally, this matches Parker et al.'s recommendation of using a minimum of 12 teams to generate an estimate of heterogeneity.) We will quantify our model's estimates and their uncertainty using Bayesian inference and will remain humble as to what we can and cannot infer from the results. We have revised the manuscript in light of these changes.

Reasons for submitting a Registered Report

We think that a Registered Report has several advantages even for a purely exploratory study: Given the large amount of degrees of freedom in assessing and modelling researchers' behavior and their research reports, we still consider a Registered Report an ideal workflow for receiving valuable expert feedback before data collection begins. This is especially important in light of the accumulated resource investment across many analysts. Moreover, all three initiating authors have strong prior beliefs that there is a substantial lack of replicability and reproducibility in our field and

that these issues can be traced back to questionable research practices. These beliefs would be confirmed by observing large flexibility in analyses and interpretation across researchers. To safeguard ourselves against our own preconceptions, preregistering this exploratory study seems adequate. Moreover, the possible implications of our results might have far-reaching consequences for psychologists working with speech, thus a Registered Report allows the research to be evaluated based solely on its theoretical rationale and its proposed methods, rather than on its outcome and interpretation.

Feasibility of collecting enough data

Beyond the minimal sample size, we intend to recruit as many analysts as possible. This justification is based on resource constraints (Lenth 2001, Lakens 2021). We aim to recruit all researchers that are qualified for the analysis, volunteering to participate, and reachable through the communication channels we will use. We consider it worthwhile collecting such a resource-constraint data set because it can inform future meta-analyses and generate worthwhile hypotheses that can inform future meta-research. Having said that, in order to gauge a rough proxy of interested parties, we circulated an interest form in which people can indicate if they would be in principle willing to join our project. This form was only circulated on Twitter (see here), thus we did not even use mailing lists, targeted emails to colleagues or other social media platforms, and still received an impressive number of responses. 216 scholars have expressed their interest. Due to privacy concerns, we cannot share the names here, but we are happy to show you the list, if you would like to see it. Even if only a third of these scholars ended up participating in the final project, we would still double or triple the sample size of Silberzahn et al. These numbers demonstrate the perceived importance of our project by scientists working on speech.

Given our arguments, our sample-size prospective analysis and the large number of interested parties in such a project, we hope we could satisfactorily address your concerns.

References:

Lakens, D. (2021, January 4). Sample Size Justification. https://doi.org/10.31234/osf.io/9d3yf

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination.The American Statistician, 55 (3), 187–193.

Parker T, Fraser H, Nakagawa S, Gould EB, Griffith S, Vesk P and Fidler F (2020) Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology [passed peer review and granted in-principle acceptance March 2020]. *BMC Biology*

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... & Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337-356.

Kind regards,

Timo Roettger (on behalf of Joseph Casillas and Stefano Coretta)