

Dear editor and reviewers,

Thank you for your detailed review of the manuscript. We have taken all the comments and suggestions into account and addressed them in the revised manuscript. First we have included additional recent literature in the introduction based on the reviewers' suggestions. Second, we have addressed comments pertaining to limited measurement reliability and taken steps to increase measurement reliability of the current study. In particular, we have decided to perform a full-null model comparison to test hypotheses. To maximise the sample size while taking into consideration limited reliability of the toddlers' preference for IDS, when testing H2a and H2b, we will weigh the contribution of the data point to the model by the number of completed trials per child and register, meaning that infants who have completed more trials will be weighted more heavily in the analysis. This will allow us to maintain a large sample while taking into consideration the limited reliability when participants don't complete the full experiment. Third, we have included questions on paternal attitudes and knowledge on language development, as well as questions regarding reading as an activity in order to distinguish between paternal caregiving and experience with reading, both of which will be used in the exploratory analyses to generate novel hypotheses for future studies. Finally, we have corrected some errors that were spotted in the first version of the manuscript, regarding trial numbers and type of auditory stimuli in the eye tracking task. Please see the following point-by-point responses to the reviewer's comments and concerns, as well as the highlighted text in the manuscript for edits and added text.

We believe that we have addressed the concerns and issues raised by the reviewers and that it has resulted in an improved manuscript. We would like to thank you for your time and feedback, and we are looking forward to hearing from you.

Yours sincerely,
The authors

Reviewer comments:

Reviewer 2:

1) The paragraph beginning «Still, the characteristics...» is confusing. It starts out stating that there is variability across cultures but then discusses Norwegian IDS in particular without a comparative lens. Then appears to return to a comparative lens but with quite narrow focus.

- Thank you for bringing this to our attention. Please see page 6 for the revised paragraph.

We have made this paragraph clearer by discussing how different characteristics of IDS varies across languages, mainly discussing Norwegian as compared to other languages. We have made the paragraph more nuanced by adding literature on differences in VSA across languages (e.g., hyperarticulation, hypoarticulation, and no difference between IDS and ADS).

2) In general, I found the review of paternal IDS a bit lean. For example, there are some older studies on e.g., the Father-Bridge hypothesis (see work by Tomasello, Berko Gleason) that might provide some relevant bigger-picture theoretical meat.

- Thank you for the comment and suggestions. Based on your suggestions, we have included more substantial literature on paternal IDS in the introduction. Please see pages 6 to 9 for more details.

3) With respect to the work on preference for IDS, work by Newman may be worth including.

- Thank you for the suggestion. We have included references to Newman's work in the revised manuscript, see page 12.

4) The claim on page 7 that "it is unknown whether fathers modulate their IDS" is too strong given that there is indeed existing literature on this topic.

- Thank you for the comment. This claim has been removed in the revised manuscript.

5) I believe there IS some literature on the impact of caregiver experience that could be explicitly mentioned (...)

- Thank you for bringing this to our attention. We have found a study by Weirich and Simpson (2019) and included this in the revised manuscript, see page 10.

6) Page 8: "The infants... will only differ in..." this is again too strong of a claim.

- Thank you for bringing this to our attention. The claim has been removed from the revised manuscript.

7) (...) What will the timeframe be (and how might it vary across fathers) between when the father completes the online questionnaire and when they come to the lab? Can the authors clarify "main caregiver" (both for review purposes and to ensure that the question is interpreted consistently by the fathers)?

- Thank you for the questions. We have now added more information, see pages 18-19 (time line) and page 13 (main caregiver), which will hopefully answer both questions.

8) Perhaps a copy of the actual questionnaire would be helpful?

- Thank you for the comment. The questionnaire (in English and Norwegian) is available at the OSF in the folder "Materials", link to the OSF page is

https://osf.io/5qjuk/?view_only=af30057f71474783a6d7629b985fa4b1

9) Perhaps the fathers who are more comfortable around their infants are more likely to take longer leaves AND produce stronger IDS?

- Thank you for bringing this to our attention. Since being comfortable around children may come from previous experience with kids, we have decided to add a question in the questionnaire and ask whether fathers have any previous working experience with kids (e.g., as a teacher in kindergarten and in school). If they have any previous experience, they are excluded from the study. Also, we have adopted the questionnaire to include questions about paternal attitudes and knowledge on language development. This would also address your concern, as fathers who think speaking in IDS is important would be more likely to speak IDS and perhaps be more comfortable to speak it to their infants. Both reading as an activity and paternal attitudes will be explored in an exploratory analysis.

10) What happens if an infant fails to calibrate?

- Thank you for the question. We see that we have been unclear on the consequences of an unsuccessful calibration. We have added this as an exclusion criterion (see page 24), meaning that infants with an unsuccessful or incomplete calibration will be excluded from the study.

11) The decision to use word lists rather than utterances is unusual (and differs from the ManyBabies study).

- Thank you for bringing this to our attention. This was an error, and it should of course have been utterances. It has thus been edited in the revised manuscript. Please see pages 20-21.

12) Why only 8 trials?

- Thank you for bringing this to our attention. This was a typo, and it should have been 16 trials (8 trial pairs). It has thus been edited in the revised manuscript. Please see page 20.

13) p.18 The comment at the bottom of the page could use some further fleshing out (how will they be transformed? How will “normally distributed” be assessed?) and this information might be better located where the other transformations are outlined, on the following page.

- Thank you for the comment. As the log transformation and deviance from normality was explained on the following page, we have removed this comment from the bottom of p.18.

Reviewer 1:

1) Perhaps the researchers could add a questionnaire to test paternal knowledge, attitudes, and/or beliefs to distinguish between these two interpretations?

- Thank you for your comment. We have edited the questionnaire to include questions about belief, knowledge and attitudes on language development in order to distinguish between paternal experience and paternal knowledge/attitudes. Please see page 19 for further description of the added questions. This measure will be added to the exploratory analyses.

2) I am a bit concerned about the choice of READING as an activity to elicit IDS. (...) Do the authors have a way of controlling how much READING fathers do with their babies and/or controlling for things like reading disorders?

- Thank you for your comment and your concern. Reading as an activity was chosen based on experience from previous studies, showing that most Norwegian parents do read to their infants and young children, and that reading does elicit differences between IDS and ADS for most acoustic features (see Rosslund et al., 2022a). In the questionnaire, there is also a question on how often the father has read to their infant the last two weeks, and this information will be reported. Fathers who do not read to their infants will be excluded from the study. We will not control for reading disorders in the current study. Please see file *Questionnaire_English_revised_after_stage1.pdf* in the folder “Materials” for the English version of the questionnaire on OSF:

https://osf.io/5qjuk/?view_only=af30057f71474783a6d7629b985fa4b1

Reviewer 3:

1) In reading the introduction several times I worried about birth order effects, but was reassured when the methods specified that all children would be first-born. It might be useful to mention this design detail earlier in the paper.

- Thank you for the comment. We have included this information earlier in the paper, please see page 5 and page 13.

2) However, one issue that has not been considered here is the measurement reliability of the infant task – how stable are individual differences as measured by the IDS-ADS preference task?

- Thank you so much for the comment. We have taken several steps to increase measurement reliability in line with your suggested research paper. First, we will compute and report the Intraclass Correlation Coefficient. Second, we will weight the contribution of the number of completed trials per child and register data points to a full-null model comparison (see Planned statistical analyses for further description). This approach will allow us to maintain a large sample size while weighting infants who contribute to more reliable data (more trials). As such, the weighted regression will take into consideration the limited reliability when the

infants don't complete the full experiment. Third, we will only include infants who completed at least half of the trials (4 in each register). And fourth, we will clarify that when drawing interference, we will not interpret non-significant correlations, meaning that absence of correlation will not be interpreted in the current study. Please see page 17 for more information about measurement reliability of the current study.