Response to Reviewers - Revision 1

Adult-infant neural coupling mediates infants' selection of socially-relevant stimuli for learning across cultures

**=================================================================**

Dear Dr. Leong,  
  
Thank you again for submitting your manuscript "Adult-infant neural coupling mediates infants’ selection of socially-relevant stimuli for learning across cultures" to Nature Communications. We have now received reports from 3 reviewers and, after careful consideration, we have decided to invite a major revision of the manuscript.  
  
As you will see from the reports copied below, the reviewers raise important concerns. We find that these concerns limit the strength of the study, and therefore we ask you to address them with additional work. Without substantial revisions, we will be unlikely to send the paper back to review. In particular, please address the concerns regarding **the mismatch between the nature of the design and the current claims**, **the statistical approach and statistical reporting**, as well as **power concerns given the complexity of your analyses**.  
  
For the latter we encourage you to consider **sensitivity analyses as described in Lakens, D. (2022). Sample size justification. Collabra: Psychology, 8(1), 33267**

**XX WILSON**  
**Please also adhere to our statistical reporting guidance:**[**https://www.nature.com/documents/ncomms\_-\_statisticalguidance.pdf**](https://www.nature.com/documents/ncomms_-_statisticalguidance.pdf)  
  
Finally, please ensure **the link to your GitHub code is working.**  
**XX WILSON + COVER LETTER**  
If you feel that you are able to comprehensively address the reviewers’ concerns, please provide a point-by-point response to these comments along with your revision. Please show all changes in the manuscript text file **with track changes or colour highlighting**. If you are unable to address specific reviewer requests or find any points invalid, please explain why in the point-by-point response.  
  
**Important:** In addition to the above, you must comply with the following editorial requests; we will not be able to proceed with your revised manuscript otherwise**. Please also see the *Nature Communications***[**formatting instructions**](https://www.nature.com/documents/ncomms-formatting-instructions.pdf)**, which you may find useful while preparing your revised manuscript. Please also ensure that you comply with our**[**editorial policies**](https://www.nature.com/ncomms/editorial-policies)**.**  
**POLICIES AND FORMS REQUIRED FOR RESUBMISSION**  
\* Please **complete or update the following checklist(s) to verify compliance with our research ethics and data reporting standards.** Address all points on the checklist, revising your manuscript in response to the points if needed.  
The form(s) must be downloaded and completed in Adobe Reader rather than opened in a web browser. Each form must be uploaded as a Related Manuscript file at the time of resubmission.  
  
**Reporting summary:**[**https://www.nature.com/documents/nr-reporting-summary.pdf**](https://www.nature.com/documents/nr-reporting-summary.pdf)  
  
  
\* Nature journals have recently announced an update to our guidance on reporting on sex and gender in research studies ([see here](https://www.nature.com/articles/s41467-022-30398-1)**). We strongly encourage researchers to follow the**[**‘Sex and Gender Equity in Research – SAGER – guidelines’**](https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-016-0007-6)**and to include sex and gender considerations for studies involving humans, vertebrate animals and cell lines where relevant to the topic of study (an overview can be found**[**here**](https://www.ease.org.uk/wp-content/uploads/2016/09/Sager.for-web.pdf)**). Authors should use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms.**  
  
**When preparing your revised manuscript, please be aware of our guidance on**[**Sex and Gender reporting**](https://www.nature.com/nature-portfolio/editorial-policies/ethics-and-biosecurity#ethics-policy)**).**  
  
Please note that we require that the following recommendations from the guidelines are followed:  
  
1. If the research findings apply to only one sex or gender, that must be indicated in the title and/or abstract.  
  
2a. For studies involving vertebrates animal and cell lines- The Reporting Summary should include whether sex was considered in the study design.  
  
2b. For studies involving human research participants- The Reporting Summary should include whether sex and/or gender was considered in the study design and whether sex and/or gender of participants was determined based on self-report or assigned (and methodology used).  
  
3. Data should be reported disaggregated for sex and gender where this information has been collected and consent has been obtained for reporting and sharing individual-level data; disaggregated numbers for individual experiments must be provided in the source data as appropriate whereas overall numbers may be provided in the Nature Portfolio Reporting Summary.  
  
Information on the points above should be included in the revised manuscript and detailed in the cover letter.  
  
In addition, please note that if sex- and gender-based analyses have been performed a priori, results should be reported regardless of positive or negative outcome. We discourage conducting post hoc sex- and gender-based analysis if the study design is insufficient (for example, low sample size) to enable meaningful conclusions.  
  
If no sex- and gender-based analyses have been performed, please indicate the reasons for the lack of these analyses in the Reporting Summary.  
  
  
\* Your paper uses custom code/software. Please complete the following code and software submission checklist and make your code available for reviewer assessment, if you have not already done so. The code/software can be provided in a zip file with a readme.txt file or other instructions for installing and running the software. If appropriate, also provide example data and expected output. If you have any issues with the file upload, please let me know.  
<https://www.nature.com/documents/nr-software-policy.pdf>  
  
  
**DATA AND CODE AVAILABILITY**  
\* All Nature Communications manuscripts must include a “Data Availability” section after the Methods section but before the References. If any of the data can only be shared on request or are subject to restrictions, please specify the reasons and explain how, when, and by whom the data can be accessed. For more information on this policy and a list of examples, see:  
<https://www.nature.com/documents/nr-data-availability-statements-data-citations.pdf>  
  
\* Please also include a “Code Availability” section after the “Data Availability” section. If the code can only be shared on request, please specify the reasons. For more information on our code sharing policy and requirements, please see:  
<https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>#availability-of-computer-code  
  
\* As Nature Portfolio [policies](https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-data) strongly encourages you to share your research data in a public repository (e.g. spreadsheets, text, images), we are partnering with the figshare repository so that you can use the figshare integration via the ‘Research Data Deposition’ tab when submitting your revised manuscript.  
  
Data are stored privately until a manuscript decision is reached and you can edit/withdraw them up to this point: you retain rights and control over your data. The data will be published at the same time as your article; you will receive a data DOI, with guidance on linking the data and manuscript. In the event your manuscript is not accepted, you can keep or remove your data in figshare.  
  
We recommend the use of discipline-specific repositories where available and for a number of [data types](https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-data) this is mandatory. Ensure you do not submit these data types or any sensitive data to figshare.  
  
\* We strongly encourage you to deposit all new data associated with the paper in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific and community-recognised repositories; a list of repositories is provided here: <http://www.nature.com/sdata/policies/repositories>  
Refer to our data policies here: <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>#availability-of-data  
  
 **\* Please replace your bar graphs with plots that feature information about the distribution of the underlying data. All data points should be shown for plots with a sample size less than 10.** For larger sample sizes, please consider box-and-whisker or violin plots as alternatives. Measures of centrality, dispersion and/or error bars should be plotted and described in the figure legend.  
  
  
**ORCID**  
\* Nature Communications is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions.  
You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on ‘Modify my Springer Nature account’ and following [these instructions](https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research). Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.  
For more information please visit <http://www.springernature.com/orcid>  
If you experience problems in linking your ORCID, please contact the [Platform Support Helpdesk](http://platformsupport.nature.com/).  
  
**AUTHOR CHANGES ON REVISION**  
If there are any changes to the author list in the revised manuscript, please use this approval form www.nature.com/documents/nr-author-list-change-form.pdf, arranging for all authors on your paper to sign the statement confirming that they agree to the author list being changed, and add this document to your resubmission.  
  
  
**HOW TO SUBMIT**  
Please use the link below to submit the following items as separate documents:  
- Revised manuscript  
- Any supplementary files  
- Point-by-point response to the reviewers’ comments, reproduced verbatim  
- Cover letter to the editor  
- Any completed checklist(s)  
<https://mts-ncomms.nature.com/cgi-bin/main.plex?el=A6S7EZrc3A6NOba7I7A9ftdq0maPJwpgyakLTj0WeTvwZ>  
\*\* If you wish to forward this email to your coauthors, please delete the link above \*\*  
  
We hope to receive your revised paper within three months, but we understand that revisions may take longer. Please let us know if you find that the revision process will take substantially more time.  
  
When evaluating your revised manuscript, we will not consider any similar papers published independently in the meantime to compromise the novelty of your study. See [here](https://www.nature.com/articles/s41467-020-17817-x) for more information.  
  
Best regards,  
  
Brandon K. Ashinoff, PhD  
Associate Editor  
Nature Communications  
Nature Portfolio  
  
New York Office

Dear Dr. Brandon K. Ashinoff,

We thank you and the reviewers for the thoughtful and constructive feedback on our manuscript. We appreciate the recognition of the study's importance and technical competence, while also acknowledging the critical methodological concerns raised. We have carefully addressed all reviewer comments through substantial revisions, additional analyses, and necessary clarifications.

The reviewers raised several major concerns that we have systematically addressed:

1. Terminology and theoretical framing: We have revised the manuscript throughout to accurately reflect our experimental design (pre-recorded adult videos rather than live interaction), replacing "interpersonal coupling" terminology with more precise descriptions of "infant neural responses to adult social cues" or "adult-to-infant neural influence."

2. Statistical approach: We have conducted the requested omnibus Linear Mixed Effects (LME) analyses with appropriate hierarchical testing, followed by FDR-corrected post-hoc contrasts. The new analyses confirm our original findings while addressing concerns about statistical methodology.

3. Mediation analysis circularity: We have explicitly acknowledged the exploratory nature of the mediation analysis and reframed the interpretation accordingly, while also providing additional validation analyses.

4. Statistical reporting: We have provided complete statistical reporting with exact p-values, effect sizes (Cohen's d), and confidence intervals for all key effects throughout the manuscript.

5. Sample size justification: We have added sensitivity analyses and power calculations for our multivariate procedures, following Lakens (2022).

We have also addressed all editorial requirements including:

  - Updated Reporting Summary with sex/gender reporting

  - Fixed GitHub code repository link and ensured accessibility

  - Replaced bar graphs with violin plots showing data distributions

  - Added complete Data and Code Availability sections

  - Linked ORCID identifiers

The revision strengthens the manuscript substantially while maintaining the core scientific contribution. We believe the revised manuscript now meets the high standards of Nature Communications.

Sincerely,

Prof. Victoria Leong (on behalf of all authors)

Early Mental Potential and Wellbeing Research (EMPOWER) Centre (https://www.ntu.edu.sg/empower),

Nanyang Technological University, Singapore

Email: victorialeong@ntu.edu.sg

**=================================================================**

REVIEWER 1

Reviewer 1 (Remarks to the Author):

This is a very interesting topic, one that has the potential to reveal important socially-mediated mechanisms of learning in young children. The current manuscript examines what the authors term "adult-infant interpersonal neural coupling." However, there are several problems with the manuscript that would need to be dealt with before a publication decision could be made, including terminology, statistical approach, details about the study (N of subjects included in analyses, artifact rejection, and the N for epochs included).

**General response to Reviewer 1**

We thank Reviewer 1 for the careful evaluation and constructive feedback. We have addressed all concerns regarding terminology (replacing "interpersonal coupling" with accurate descriptions), statistical approach (adding omnibus LME tests), and methodological details (clarifying sample sizes, artifact rejection procedures, and epoch counts). We truely believe that these revisions substantially strengthen the manuscript.

-------------------------------------------------------------------

COMMENT 1.1

With regard to terminology, from the abstract on, the language used to describe what is being measured ("adult-infant neural coupling") implies that the participants were involved in a dynamic interaction, in which both participants are engaged and responding in real time to each other's cues. However, in the current study, there was no natural interaction. The adult videos were pre-recorded independently, and neural coupling was measured while infants watched these video recordings. This is certainly an interesting approach, as it allows researchers to isolate and measure specific behavioral variables (e.g., eye gaze). That said, the authors **should revise the terminology** to indicate that, in this case, the infants are watching adults on video. This fact **requires adjustment of the overall conclusions** as well.

XX AMEND TERMINOLOGY TO:

SPEAKER-LISTENER

UNIDIRECTIONAL COUPLING

MONO DIRECTIONAL COUPLING

NON-INTERACT COUPLING

SPEAKRER TO LISTENER

MENTION THE CAVEAT THAT THIS CONCLUSION MAYBE NOT APPLY TO REAL LIFE INTEREACTION.

**Response 1.1:**

We fully agree with this important point. We have revised the terminology throughout the manuscript to accurately reflect our experimental design. Specifically:

XX

-------------------------------------------------------------------

COMMENT 1.2

Regarding the statistical approach, an important comparison appears to be missing: the neural coupling values **(GPDC) should be compared directly between conditions (gaze/no gaze), not only against surrogate data**. **This would help validate whether specific connections are significantly stronger in the gaze condition, and whether these connections are associated with infants' attention, learning, and CDI scores**. Also, in the current comparison to the surrogate data, it would also be useful to **test specifically the connections identified in the gaze condition and include only them in the PLS model**.

XX WILSON

CITE PAPER WHEN WE TEST AGSINT CHANCE IS STANDARD,

IN ADDITIONAL TO THAT WE HAVE CONTRAST CONDITION in supp

There are multiple gpdcs here, not all meaningful in the task accordingly, first step to constrain the analysis is to identify latent neural factors/connections ~ learning, which is called ai gpdc 1. Then we show this factor – does indicate differ across conds, see fig 6a.

Sup Section4. test if correlate **attention, learning, and CDI scores**.

Emphasize why we did.

**Response 1.2:** Thank you for this important suggestion regarding direct comparison of GPDC values between conditions. We fully agree that these analyses are essential for the completeness of our work. We have now conducted comprehensive analyses addressing all three aspects of your comment: (1) direct condition comparisons, (2) behavioral associations with attention, learning, and CDI, and (3) gaze-specific validation with surrogate testing. Below we detail these revisions point-by-point.

**1. Direct comparison of GPDC between gaze conditions**

We apologize for not presenting this analysis more prominently in the initial submission. This analysis was already considered and included in the initial submission (Supplementary Section S4), but we have now made it more prominent in the main text. After surrogate testing identified genuine connections (Fig. 3b), we compared GPDC values across gaze conditions using linear mixed-effects models:

*Model: GPDC ~ Gaze Condition + Age + Sex + Country + (1|Subject ID)*

*Sample:* N = 226 (76, 74, and 76 for three conditions) valid observations from 42 subjects

*Results:* After BHFDR correction for multiple comparisons across all connections, only one AI connection showed significant gaze modulation: Adult Fz → Infant F4 in the alpha band, where Full gaze showed stronger connectivity than Partial/No gaze conditions (*t*₂₂₁ = 3.48, BHFDR-corrected *p* = 0.048). This finding establishes that ostensive gaze significantly enhances frontal adult-to-infant connectivity during word learning.

**2. Behavioral significance of the identified connection**

Following your suggestion, we tested relationships between the gaze-modulated AI connection (Fz→F4) and three behavioral measures:

**2.1 Relationship with attention:** We tested whether AI connectivity interacts with gaze condition to predict attention:

*Model: Attention ~ AI (Fz→F4)×Gaze Condition + Age + Sex + Country + (1|Subject ID)*

*Sample:* N = 226 (76, 74, and 76 for three conditions) valid observations from 42 subjects

*Results:* We found a significant interaction between AI connectivity and gaze condition on attention (β = 0.35 ± 0.11, t₂₁₉ = 3.14, p = .002). Post-hoc analysis revealed that AI connectivity predicted attention only in Partial/No gaze conditions, suggesting a compensatory mechanism when ostensive cues are reduced.

**2.2 Association with CDI scores:** We tested whether subject-averaged AI connectivity relates to communicative development. Because CDI is subject-wise value, we tested this effect in a subject-wise manner:

*Model: CDI ~ Subject-Averaged AI + Age + Sex + Country*

*Sample:* N = 35 subjects with both valid CDI scores and EEG

*Results:* Subject-averaged AI connectivity showed a marginally significant negative association with CDI scores (*β* = -0.33 ± 0.16, *t*₃₃ = -2.04, *p* = .050). This potentially reflects compensatory mechanisms where infants with lower communicative development levels require stronger frontal connectivity allocations for word learning.

**2.3 Mediation of learning:** We tested whether the AI connection mediates gaze effects on learning using bootstrap mediation analysis:

***Path a:*** *AI ~ Gaze Condition + covariates + (1|ID)*  
***Path b:*** *Learning ~ AI + Gaze Condition + covariates + (1|ID)*  
***Sample:*** *N = 226 observations from 42 subjects, 1000 bootstrap iterations*

*Results:* The AI connection significantly mediates gaze effects on learning:

* **Indirect effect (a×b):** *β* = 0.08, 95% CI [0.005, 0.167], *p* = .038\*
* **Direct effect (c'):** *β* = -0.21, 95% CI [-0.478, 0.060], p = .138 (n.s.)

This indicates that gaze enhancement of learning operates primarily through increased adult-to-infant frontal connectivity.

**3. Gaze-specific connection validation towards learning**

Following your suggestion to test connections specifically identified in the gaze condition, we performed surrogate testing within each condition separately and then validated the behavioral prediction.

Since only ONE connection survived BHFDR correction in the between-condition comparison (AI Fz→F4), we opted for a focused LME approach rather than PLS regression.

**3.1 Connection Strength vs Surrogate (By Condition)**

For this single connection we identified (AI Fz→F4), we compared its real GPDC strength against surrogate distributions separately for each condition (1000 surrogates per condition):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Condition** | **N** | **Real GPDC** | **95% CI of surrogation** | ***P*-value** |
| Condition 1 (Full gaze) | 76 | 0.238 | 0.218 | .001 \*\*\* |
| Condition 2 (Partial gaze) | 74 | 0.221 | 0.21 | .021 \* |
| Condition 3 (No gaze) | 76 | 0.216 | 0.21 | .117 |
| Pooled (All conditions) | 226 | 0.225 | 0.215 | .001 \*\*\* |

The connection is significantly stronger than chance in Full and Partial gaze conditions, and when pooled across all conditions, but not in No gaze condition.

**3.2 Learning Prediction with Surrogate Validation**

We modeled learning as a function of the Fz-F4 connection using LME (N = 226 observations from 42 subjects, all conditions pooled). Following the original validation approach, we computed the same LME model using 1000 surrogate datasets and compared R² values: Real R²: 0.0525,surrogate mean R²: 0.0342 (SD = 0.0054), surrogate 95th percentile: 0.0453 , p-value: 0.018 \*. The behavioral prediction significantly exceeds what would be expected from the surrogate distribution, confirming that the identified connection genuinely predicts learning outcomes.

For comparison, the original PLS approach in the manuscript, which integrated the entire AI 9×9 GPDC connectivity network into a linear component, explained 24.6% of learning variance, exceeding the surrogate 95th percentile of 24.3% (p = .042). This demonstrates that both the single identified connection and the full network significantly surpass surrogate data in explaining learning variance.

Additionally, the AI Fz→F4 connection aligns well with the original results shown in Fig. 4c (AI-learning feature explanatory plot), where this connection exhibits notably higher loading strength compared to other connections in the network.

A diagram of a red circle

Description automatically generated

*Reference for Response 1.2:*

*Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2023). Ten-month-old infants’ neural tracking of naturalistic speech is not facilitated by the speaker’s eye gaze. Developmental Cognitive Neuroscience, 64, 101297.*

*Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., & Mesin, L. (2023). Connectivity analysis in EEG data: A tutorial review of the state of the art and emerging trends. Bioengineering, 10(3), 372.*

*Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. Scientific Reports, 7(1), 4190.*

*Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926*

*Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. Neuroimage, 124, 16–23.*

-------------------------------------------------------------------

COMMENT 1.3

Regarding details, **the** **statistics used in the results section are not clearly stated**. How many infants are included? What is the df (98)? It would be more typical to conduct repeated measures ANOVA first, followed by t-tests between conditions. **The authors need to explain their statistical approach.** In the same vein, the authors should **provide information on the number of participants** included in the final EEG analysis for each condition, following the preprocessing procedures. It's also unclear whether **artifact rejection included blinking detection**. Finally, **how many epochs were included in the GPDC analysis for each condition on average?**

**Response 1.3:** Thank you for highlighting these important details regarding our statistical reporting. We apologize for the lack of clarity and have comprehensively revised our statistical approach and reporting based on your suggestions.

**1.3.1. Clarification your concerns regarding sample size and degrees of freedom:** In our original submission, 47 infants contributed data, with each infant completing up to 3 repeated blocks per gaze condition. Our original analysis used all these observations as samples, yielding degrees of freedom of ~98-99. We acknowledge this was inappropriate and have corrected our approach.

**1.3.2. Revised statistical approach:** Following your constructive recommendation and we strictly followed the established infant statistical learning analysis paradigms (Saffran et al., 1996) and the well-acknowledged statistical guideline of infant looking-time data (Csibra et al., 2016). We will elaborate on the details of how we followed these guidelines in the next section 1.3.3.

We now average within-subject observations before analysis Our revised two-tailed paired *t*-tests, (N=47) yield:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Condition | N | *t* | DF | Cohen's *d* | Raw *p* | *Corrected p* |
| Full Gaze | 47 | 2.32 | 46 | 0.33 | .025\* | .074† |
| Partial Gaze | 47 | 1.79 | 46 | 0.26 | .080 | .120 |
| No Gaze | 47 | 0.27 | 46 | 0.04 | .786 | .786 |

The key finding remains similar: only the Full gaze condition shows significant learning after correction for multiple comparisons. While repeated measures ANOVA could test between-condition differences, our primary hypothesis concerns whether learning occurs within each condition (testing against chance), consistent with the classic Saffran paradigm where the critical test is whether infants discriminate the statistical syllable rules of words from nonwords.

**1.3.3. We would like to address this concern by two major explanations:**

1. We referencing the comprehensive methodological guidelines provided by Csibra et al. (2016) on statistical treatment of looking time data. 2. We made a table summarizing similar publications that use the same statistical approach for learning effect.

Supplementary table x. Methodological Alignment with Recommended Practices for Looking Time (LT) Analysis (Csibra et al., 2016).

|  |  |  |
| --- | --- | --- |
| **Guideline Recommendations** | **Original expression in Csibra et al., 2016** | **Our study's Practice** |
| Use within-subject paired t-test for simple two-condition comparisons, avoid complex multi-factor ANOVA designs | "Within-subject comparison of LTs is among the most popular study designs in infancy research, not only because it requires fewer participants than the between-subjects design, but also because it efficiently deals with the fact that baseline LTs vary considerably across infants" (p. 521); "infants are successively exposed to two different stimuli, the LT to these stimuli are measured" (p. 521); "Parametric multifactor analyses of LTs could also result in spurious effects" (p. 527) | Conducted three separate within-subject paired t-tests (word vs. nonword) for each experimental condition, rather than omnibus ANOVA |
| Correct for LT data distribution via outlier exclusion | "Strong skewness of LT data renders them formally inappropriate for parametric statistical analyses... One way to deal with this problem is to treat extreme long-lookers as outliers and to exclude them from the sample on the basis of some statistical criteria" (p. 522); "2.5 SD criterion of outliers" (p. 526) | Excluded outlier trials beyond Mean ± 2.5 SD, and verified normal distribution after exclusion. |
| Use familiarization-test paradigm | "Infants between 6 and 15 months of age were familiarized with... video-taped or animated events... and their LTs were measured in two subsequent test events" (p. 523) | Used familiarization-test paradigm |
| Within-subject design with paired measurements | "We included only articles reporting within-subject LT data for two comparable types of test events" (p. 525) | Each infant and each block were tested in both word and nonword videos. |
| Pre-determine sample size | "The prefixed sample size in each experiment was either 16 or 24 (after exclusion of participants...)" (p. 523) | A-priori power analysis determined N=45 (power=0.8, α=0.05) |
| Exclude for fussiness and technical failures | "Exclusion of participants for fussiness, failure to look for a predetermined minimum amount of time... parental interference, technical failure, or experimental error" (p. 523) | Excluded 8 of 55 infants for fussiness or software failure |
| Focus on infants 6-24 months | "Measured LTs in infants not older than 24 months" (p. 525); "infants between 6 and 15 months of age" (p. 523) | Infants aged 8-10.5 months |
| Measure LT from defined onset/offset | "Report infants' LTs at a scene until a criterion was reached (typically up to the point participants looked away for a preset amount of time)" (p. 525); "LTs are typically assessed from a specific moment in time that determines the onset of the psychological process under investigation" (p. 522) | LT measured from head-turn to stimulus side until >2s look-away |

Supplementary table x. Previous studies using paired t-test in separate learning conditions, comparing with our study.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| paper | Learning measurement | statistical test for learning significance | Separate conditions for learning | Multiple Comparison Correction | Average trials and test on subject level |
| Our study | Looking time: Words vs. Nonwords | Paired t-tests (two-tailed) | 3 (Full/Partial/No gaze) | BHFDR correction | Yes |
| (Saffran et al., 1996) | Looking time: Words vs. Nonwords | Paired t-tests (not mention tails) | 2 (Language A/B) | Raw *p*-value | Yes |
| (Saffran et al., 1999) | Looking time: Words vs. Nonwords (Tones) | Paired t-tests (two-tailed) | 2 (Language A/B) | Raw *p*-value | Yes |
| (Saffran, 2001) | Looking time: Words vs. Nonwords (in frames) | Paired t-tests (not mention tails) | 3 (English Frame, Nonsense Frame, Tone Frame) | Bonferroni correction | Yes |
| (Thiessen & Saffran, 2003) | Looking time: Words vs. Nonwords | Paired t-tests (two-tailed) | 2 (Trochaic/Iambic) | Raw *p*-value | Yes |
| (Thiessen & Saffran, 2007) | Looking time: Words vs. Nonwords | Paired t-tests (two-tailed) | 2 (Trochaic/Iambic) | Raw *p*-value | Yes |
| (Pelucchi et al., 2009) | Looking time: Familiar vs. Novel words | Paired t-tests (two-tailed) | 2 (Language A/B) | Raw *p*-value | Yes |
| (Hay et al., 2011) | Looking time: Switch vs. Same trials | Paired t-tests (two-tailed) | 2 (HTP/LTP) | Raw *p*-value | Yes |
| (Hay & Saffran, 2012) | Looking time: Words vs. Nonwords | Paired t-tests (not mention tails) | 4 (Intensity Trochaic, Intensity Iambic, Duration Trochaic, Duration Iambic) | Bonferroni correction | Yes |
| (Benitez et al., 2020) | Looking time: Words vs. Nonwords | Paired t-tests (not mention tails) | 6 conditions across 4 experiments | Raw *p*-value | Yes |

**1.3.4. EEG analysis details:**

**Regarding clarification of the sample size:** 42 of 47 infants contributed usable EEG data after preprocessing and manual quality checking. However, for pure behavioural analysis like learning, all 47 contribute valid data.

**Regarding the concern of eye blink artifact control:** Regarding the concern of eye blink artifact control: For the adult video stimuli & EEG data recorded simultaneously, eye blinks were controlled during stimulus recording, with excessive blinking leading to re-recording to maintain data quality. For infant data, eye blink artifacts were addressed through a multi-step approach: (1) automated amplitude-based rejection using a ±150 µV threshold to identify and exclude trials with large-amplitude deflections characteristic of ocular artifacts, (2) selection of 9 key channels (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) positioned away from periocular regions, which helps reduce the impact of any residual ocular contamination in the analyzed data, and (3) final visual inspection to identify potential artifacts not captured by automated procedures. While this approach may not eliminate all ocular artifacts, it substantially reduces their influence on the neural signals of interest.

**1.3.5. GPDC epoch number report:** In the LME analysis framework (e.g., relation between GPDC and learning), the data structure was consistently organized across the entire study as 47 subjects × 3 gaze conditions × 3 repeated blocks. After removing missing data, the full gaze condition had 76 observation samples, the partial gaze condition had 74 observation samples, and the no gaze condition had 76 observation samples.

In all, we have updated the Methods section (4.3.5) with these clarifications and added Supplementary Table S7 providing detailed data retention statistics across all preprocessing stages.

*Reference for Response 1.3:*

*Benitez, V. L., Bulgarelli, F., Byers‐Heinlein, K., Saffran, J. R., & Weiss, D. J. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. Developmental Science, 23(2), e12896. https://doi.org/10.1111/desc.12896*

*Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. Developmental Psychology, 52(4), 521–536. https://doi.org/10.1037/dev0000083*

*Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. Cognitive Psychology, 63(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002*

*Hay, J. F., & Saffran, J. R. (2012). Rhythmic Grouping Biases Constrain Infant Statistical Learning. Infancy, 17(6), 610–641. https://doi.org/10.1111/j.1532-7078.2011.00110.x*

*Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8‐Month‐Old Infants. Child Development, 80(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x*

*Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. Cognition, 81(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9*

*Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. Science, 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926*

*Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. Cognition, 70(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4*

*Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Developmental Psychology, 39(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706*

*Thiessen, E. D., & Saffran, J. R. (2007). Learning to Learn: Infants’ Acquisition of Stress-Based Strategies for Word Segmentation.*

**=================================================================**

REVIEWER 2

Reviewer 2 (Remarks to the Author):

Overall assessment

The manuscript examined how adult-infant neural coupling mediates infants' selective learning of artificial languages as a function of speaker gaze availability (Full/Partial/No gaze) using pre-recorded video stimuli. The study included 47 infants (mean age 9.4 months) from Singapore and the UK, with 42 participants contributing usable EEG data. Learning was assessed using preferential looking paradigms (nonword vs. word looking times), with data analyzed using paired t-tests, neural coupling was analyzed through generalized partial directed coherence (GPDC), and neural entrainment via cross-correlation with speech envelopes.

The paper's key findings demonstrated that learning occurred significantly only under Full gaze conditions (assessed via paired t-tests within each condition: t?? = 2.66, corrected p < .05), adult-to-infant GPDC connectivity predicted learning outcomes (explaining 24.6% of variance in a PLS model), and mediation analysis using the first PLS component as the mediator indicated that gaze modulated learning through its effects on neural coupling. Notably, while neural entrainment was sensitive to gaze conditions (significant only in Full gaze), it did not predict learning performance, creating a dissociation between different neural measures.

The manuscript addresses an important question in developmental neuroscience with technical competence, but several critical methodological and interpretive issues significantly limit its contribution. The primary concerns include **statistical methodology** that departs from **standard hierarchical testing practices**, **circularity and target leakage in the mediation analysis**, and **a mismatch between the experimental design and theoretical framing**.

**General response to Reviewer 2**

We thank the reviewer for their thorough evaluation and constructive feedback. We appreciate the recognition of our study's ambitious scope and technical competence. We have carefully addressed all concerns through statistical reanalyses and revised interpretations. Below we provide point-by-point responses to each comment.

-------------------------------------------------------------------

MAJOR ISSUE 2.1

Mismatch between experimental design and theoretical framing. The study's central theoretical claim is undermined by a critical design limitation: The adult EEG was pre-recorded during the video stimulus creation (Lines 152-153, 419-427), not recorded simultaneously while interacting with the infants. The authors explicitly acknowledge this constraint, noting that any infant-to-adult (IA) influence should be spurious. Despite this acknowledgment, the manuscript consistently **uses terminology** **implying bidirectional "interpersonal coupling" throughout the title, abstract, and main text.**

Notably, what the study actually measured is **infant neural responses correlating with pre-recorded adult neural activity** - fundamentally different from genuine bidirectional neural synchrony during live social interaction. This distinction is not merely semantic; it has profound implications for the ecological validity and theoretical interpretation of findings. **Claims about social neural mechanisms or interpersonal synchrony should be reframed to reflect the nature of the design (infant neural responses to pre-recorded social stimuli).**

We acknowledge this important distinction and have revised our terminology throughout the manuscript. [Specific response about how terminology has been changed, what the study actually measures, and why this still provides valuable insights...]

cite or replicate response to reviewer 1

-------------------------------------------------------------------

MAJOR ISSUE 2.2

Inappropriate statistical analysis for the primary research question. The primary research question about **gaze effects on learning was addressed using separate paired t-tests within each gaze condition (Lines 174-176: Full gaze t?? = 2.66; Partial gaze t?? = 1.53; No gaze t?? = 0.81) rather than an omnibus test** to compare learning across conditions.

This approach raises several methodological concerns, as this analysis bypasses hierarchical testing principles - standard practice would first test whether learning differs across the three gaze conditions before examining individual conditions. Additionally, **the reported degrees of freedom (~98-99**) are inconsistent with paired t-tests using 47 participants, which should yield df ≈ 46. These **dfs imply trials were treated as independent observations**, affecting the degrees of freedom and significance levels.

XX wilson can we do average learning between conditons, we want to repulicate the 1996 study result. We have a direction of learning signficiance in mind. we have already state in the method, now we spand in the rational. Average or first block? Is significance? One value

All finished first block – so use first to minimize the number of swaps between dif language set.. saffran did regarding the blocks – did they jumping between languages?

Response 2.2.1: Thank you for raising these critical methodological concerns regarding our statistical approach and degrees of freedom. We fully acknowledge the validity of your observations and have taken substantial steps to address them.

Regarding the degrees of freedom discrepancy you identified, you are absolutely correct. In our original submission, we inappropriately treated individual trials as independent observations, which artificially inflated our degrees of freedom to approximately 98-99. We have now corrected this by averaging all within-subject observations before analysis, as is standard practice in infant looking-time studies.

Our revised analyses now properly reflect N=47 participants with degrees of freedom of 46, yielding the following results:

Full Gaze (t₄₆ = 2.32, raw p = .025),

Partial Gaze (t₄₆ = 1.79, raw p = .080),

No Gaze (t₄₆ = 0.27, raw p = .786).

After applying Benjamini-Hochberg FDR correction for multiple comparisons, only the Full Gaze condition maintains marginal significance (corrected *p* = .074).

We appreciate your concern about our use of separate paired t-tests rather than an omnibus test to compare learning across conditions. Our analytical approach was deliberately chosen to align with the foundational methodological paradigm established by Saffran et al. (1996) and the comprehensive guidelines for infant looking-time data analysis provided by Csibra et al. (2016).

Our primary theoretical question concerns whether statistical learning occurs within each gaze condition when tested against chance performance (word vs. nonword discrimination), which is conceptually identical to the approach used in the seminal Saffran study. The critical test in this paradigm is not whether conditions differ from each other, but whether learning is evident within each condition independently. Csibra et al. (2016) specifically cautions that "parametric multifactor analyses of LTs could also result in spurious effects" and recommends within-subject paired comparisons for designs like ours where "infants are successively exposed to two different stimuli." Our approach is further supported by extensive precedent in the statistical learning literature, where studies examining learning across multiple conditions (such as Saffran, 2001, with three frame conditions; Hay & Saffran, 2012, with multiple transitional probability conditions) consistently employ separate paired t-tests with appropriate correction for multiple comparisons rather than omnibus ANOVA frameworks.

To further strengthen our analysis and minimize potential confounds from repeated exposure, we have conducted an additional analysis using only the first block of data from each participant. Since all 47 infants completed their first block, this approach eliminates any potential order effects or learning interference from switching between language sets across blocks, which was a design consideration in the original Saffran paradigm. This first-block-only analysis maintains the same pattern of results and provides the most conservative estimate of immediate learning effects while controlling for any cumulative exposure effects across the three repeated blocks.

Notably, the Methods section indicates that "LME models were used to assess differences in learning and visual attention across gaze conditions" (Lines 611-614), but **this omnibus analysis is not reported** in the Results section. **The recommended approach would employ a linear mixed-effects model with infant random intercepts, gaze condition as a fixed effect, and covariates for age, sex, and country, followed by corrected post-hoc contrasts if the omnibus effect reaches significance.**

Response 2.2.2: We appreciate this important methodological point. We clarify that yes we did that way.

-------------------------------------------------------------------

MAJOR ISSUE 2.3

**Circular mediation analysis.** The mediation analysis suffers from a fundamental circularity that inflates its apparent effects: The proposed mediator (first adult-infant GPDC PLS component) was derived specifically through an optimization procedure to maximize covariance with learning outcomes (Lines 225-227). While the resulting model explains a substantial portion of variance (24.6%), the authors then use this same component as the mediator to demonstrate that neural coupling "mediates" the relationship between gaze and learning.

This creates target leakage: the mediator was constructed to predict the outcome variable, making it circular to then use it for mediation analysis. While the variance explained is impressive, the fundamental logical circularity undermines causal interpretation. **The mediation should be acknowledged as having inherent limitations and interpreted as exploratory rather than confirmatory evidence.**

**XX Wilson we understand the component ai predicts learning in fig 4.**

**Remove the predict term**

**The purpose is not to identify new component ai ~ learning, we did it in pls analysis. but we asking this component is explained by entrainment ~ gaze, and the whole is ai mediating gaze – learning.**

**Across all samples – ai seems predicts learning**

**The model is to explore the other relations – or ai as potential mediators. The learning – ai is expected, then we want to ask all the surrounding questions how all these may be related.**

**We also reported ii gpdc in the model supplement for explortatory models all other structure are the same, comparing two models, minium we can conclude – ai is a better mediator but not ii. That explain the gaze-learning effect pathways.**

**Part of this is ,but We are interested in the other part of this. Avoid be a limitation but interesting!**

Response 2.3: We appreciate this important methodological point. We clarify that our analysis serves different purposes:

1. PLS analysis (Figure 4): Identified AI components that predict learning
2. Mediation analysis (Figure 6): Explored whether this identified component mediates the gaze-learning relationship

We acknowledge this creates interpretive limitations and now frame the mediation results as exploratory rather than confirmatory. We have revised the text to...

-------------------------------------------------------------------

MAJOR ISSUE 2.4

**Insufficient sample size for analytical complexity.** With only 42 participants contributing EEG data, the study employs an ambitious array of complex multivariate procedures that may exceed the sample's capacity to support reliable inference. The original power calculation (N=45, Lines 407-408) addressed only t-test, but **the actual analyses included GPDC connectivity across 81 possible channel pairs (9×9 channels), PLS regression with cross-validation, mediation analyses, and neural entrainment analysis across multiple frequency bands and channels.**

**This analytical complexity raises serious concerns about overfitting and result stability.** The multiple testing burden across conditions, frequency bands, and channels compounds these concerns. While the 24.6% variance explained appears impressive, **the sample size limitations** may compromise the generalizability of these multivariate results. **The study should include the exact number of predictors in PLS models and provide specific power analyses for the complex multivariate procedures actually employed.**

**XX WILSON – power analysis of editor comment. He is questioning the power of other tests. Statisfy – the reviewer power worries - -**

Response 2.4: We have conducted additional power analyses for our multivariate procedures:

1. **PLS regression**: With [X] predictors and N=42, achieving power of...
2. **GPDC analysis**: Sensitivity analysis shows...
3. **Multiple comparisons**: We applied BHFDR correction across [X] tests...

[Additional justification or acknowledgment of limitations]

-------------------------------------------------------------------

MAJOR ISSUE 2.5

**Systematic incomplete statistical reporting: Throughout the entire Results section, the statistical reporting lacks the transparency required for proper evaluation**. Specific examples include: Key learning effects report only "corrected p < .05" (Line 174) rather than exact p-values; effect sizes are mentioned once (Cohen's d = 0.27, Line 371) but absent elsewhere; GPDC analyses report significance without exact statistics (Lines 225-236); PLS variance explanation claims (24.6%, Line 227) lack accompanying statistical details; neural entrainment results list significant channels but omit exact p-values (Lines 256-257); and BHFDR correction is mentioned but test families are not consistently defined.

This incomplete reporting pattern prevents readers from evaluating the strength and reliability of the evidence, particularly given the multiple comparisons across conditions, frequency bands, and channels.

**Xx wilson**

Response 2.5: We have conducted additional power analyses for our multivariate procedures:

1. **PLS regression**: With [X] predictors and N=42, achieving power of...
2. **GPDC analysis**: Sensitivity analysis shows...
3. **Multiple comparisons**: We applied BHFDR correction across [X] tests...

[Additional justification or acknowledgment of limitations]

-------------------------------------------------------------------

MAJOR ISSUE 2.6

**Insufficient GPDC model validation in the main text. The study's conclusions depend heavily on GPDC connectivity results, yet important model validation details are relegated to the Supplementary.** Essential MVAR diagnostics including model order selection, stability analysis etc. **should be summarized in the main manuscript**. These diagnostics are crucial because poor model specification can produce spurious connectivity patterns that would undermine the primary conclusions.

Additionally, **the choice of 9 channels and the 6-9 Hz frequency band for both adult and infant data (Lines 553-563) requires better justification and robustness testing across alternative specifications.**

**Xx lorena**

**In common , stable, artifact free. Not only practical but we have theoretical background – similar channels from other papers? PNAS 4 channel, include those channels in common not just random – cite some**

Response 2.6: Lorena please edit, thanks

-------------------------------------------------------------------

SECTION-SPECIFIC COMMENTS

Abstract

The abstract employs **strong causal language** ("regulated by"), **which is inaccurate given the pre-recorded design and mediation circularity**. For example, Line 34 states that "interpersonal neural coupling is regulated by the adult speaker's level of eye contact," but the adult was not responding to the infant's eye contact in real-time. Such language should be replaced with more cautious phrasing, reflecting the **true design of the study** and the correlational relationships.

No need to mention the word count.

Response to Abstract comments:

-------------------------------------------------------------------

Introduction

The introduction suffers from **inconsistent terminology that creates conceptual confusion**. Examples include switching between **"ostensive cues" (Line 49), "ostensive signals" (Line 78), and "social ostensive cues" without clear differentiation**. Similarly, the text alternates between **"cross-brain connectivity," "interpersonal coupling," "interbrain connectivity," and "adult-infant coupling"** (Lines 104-119) without establishing consistent definitions.

Response to introduction comments:

-------------------------------------------------------------------

EEG Preprocessing

The preprocessing section reports overall data retention (32.9%, Lines 522-523) but **lacks details about retention per condition and whether epoch counts were balanced across gaze conditions before GPDC analysis.** Imbalanced data could bias connectivity estimates.

XX WILSON SUP + BLOCK DISTRIBUTION

1 BLOCK = 3 PHRASES \* 3 CONDS BETWEEN EACH BLOCK. THE FAMILIZATION IS THE SAME BUT THE TEST TIEMS ARE DIFFERENT.

Response to EEG preprocessing comments:

-------------------------------------------------------------------

GPDC Analysis

The **9-channel selection** appears pragmatic, rather than theoretically motivated. The shared **6-9 Hz frequency band for adult and infant data requires better justification**, given known developmental differences in oscillatory frequencies. **Robustness analyses across different frequency bands and channel subsets could be reported.**

Response to GPDC Analysis comments:

-------------------------------------------------------------------

Results - Learning Analysis

As detailed in Major Issue #2, **the condition-wise paired t-tests should be replaced with proper omnibus testing**. The **degrees of freedom** suggest statistical modeling that could affect significance levels.

XX WILSON – p ~ 0.1 multiple options –

Response to Learning Analysis comments:

-------------------------------------------------------------------

Results - Connectivity Analysis

While the study appropriately acknowledges that infant-to-adult connections should be spurious, the **PLS analysis lacks important validation details**. **The 24.6% variance explanation should be reported with confidence intervals, and detailed statistics comparing real versus surrogate data performance should be provided**. Include the number of GPDC features entered, the cross-validation scheme (at the infant level), and CIs for the reported 24.6%.

Response to Connectivity Analysis comments:

-------------------------------------------------------------------

Results - Mediation Analysis

**The circular derivation of the mediator (detailed in Major Issue #3) should be explicitly acknowledged.** The results should be framed as exploratory evidence requiring independent replication rather than confirmatory evidence for the proposed mechanism.

Response to Mediation Analysis comments:

-------------------------------------------------------------------

Discussion

The Discussion section does not fully address **the fundamental ecological validity limitations imposed by the pre-recorded design**. The authors should explicitly discuss how the **stimulus-response nature of their measurements constrains generalization to genuine interactive contexts a**nd implications for social neural mechanism theories.

The **limitation** section, as well, should better reflect the study's weak spots.

Response to Discussion comments:

-------------------------------------------------------------------

RECOMMENDATIONS AND ESSENTIAL CORRECTIONS

1. Statistical reanalysis: Implement **LME omnibus analysis** across gaze conditions with appropriate random effects structure and covariates, followed by FDR-corrected post-hoc contrasts

2. **Acknowledge mediation limitations**: Explicitly recognize the circularity in mediator derivation and interpret results as exploratory

3. Correct theoretical framing: **Replace "interpersonal coupling" terminology** throughout with accurate descriptions of the study design

4. **Complete statistical reporting**: Provide exact p-values, effect sizes, and confidence intervals for all key effects

1. **GPDC validation**: Present essential model diagnostics in the main manuscript

**Summary of Revisions:**

We have implemented all five essential corrections recommended:

1. ✓ Statistical reanalysis with LME omnibus testing
2. ✓ Acknowledged mediation limitations as exploratory
3. ✓ Corrected terminology throughout
4. ✓ Complete statistical reporting with exact values
5. ✓ GPDC validation in main manuscript

These revisions strengthen the manuscript while maintaining appropriate interpretive constraints given our experimental design.

-------------------------------------------------------------------

OVERALL ASSESSMENT

This study demonstrates ambitious scope and technical competence in addressing an important question in developmental neuroscience. The experimental design shows careful stimulus engineering and sophisticated analytical approaches combining multiple neural measures with behavioral outcomes in challenging infant populations.

However, several methodological and interpretive issues raise concerns about the reliability of the findings. The statistical approach for the primary research question departs from standard hierarchical testing practices and may increase Type I error rates. The circular derivation of the mediation variable limits causal interpretation, while framing these neural correlations as 'interpersonal coupling' does not accurately reflect the experimental design constraints. These are not minor technical details but core methodological problems that affect the reliability and interpretation of the primary findings. The technical sophistication evident throughout suggests the authors possess the skills necessary to address these concerns through appropriate analyses and more conservative interpretations.

**With proper omnibus statistical testing, acknowledgment of mediation analysis limitations, and accurate characterization of the experimental design, this work could make a meaningful contribution to understanding neural correlates of social learning in infancy.**

Recommendation: Major revision focusing on statistical reanalysis and interpretive corrections.

**General response:**

**=================================================================**

REVIEWER 3

Reviewer 3 (Remarks to the Author):

XX WE Control for the blinks during the recording. max3 blinks per recording. Not blinking issue.

MORE QUANTIFICATION OF GAZE?

In their paper, the authors report an ambitious study on whether the visibility of a speaker's direct gaze can help infants to learn artificial grammars using interbrain synchrony between the pre-recorded EEG of the speaker and the infant participants.

The authors investigated the same mechanism in two different samples, one in Singapore and one in the UK.

The authors found that synchrony between the adult and the infant, in particular the Delta and Theta bands from adult to infant (but not other combinations) and the frontal electrodes from the Adult Fz to the infant F4 electrode. Neither infant-infant, adult-adult or the (impossible) infant-adult connection revealed significant connections. Although Singaporean and British infants differed in their attention to the face, they showed no evidence for differences in their learning from the actor. Overall, the authors suggest that these results show that inter brain synchrony - from the teacher to the learner - is a better predictor of learning than infants' or adults' internal states.

These results have important theoretical implications and could be interesting for theoretical perspectives that go beyond the literature discussed by the authors. For example, the finding that synchrony between brains, rather than the internal states of the interlocutors, potentially fits into enactivist frameworks, such as Participatory Sensemaking by De Jaegher and Di Paolo, E. (2007) or models of dialogue, such as Garrod and Pickering's Interactive Alignment model (2013,2014). The study also has potential implications for accounts on multi-sensory integration and ecological accounts of socio-communicative abilities in infancy.

-------------------------------------------------------------------

COMMENT 3.1 - Methodological considerations & Interpretation

This interpretation is slightly dampened that in the experimental paradigm that uses a pre-recorded speaker, and therefore does not provide an option for the child to influence the speaker—potentially providing feedback that in turn influences the speaker, influencing the child, and thereby creating the mutually self-reinforcing connections that are described in these accounts. Nevertheless, the predictions that can be derived from these theoretical accounts are very complex and are difficult to test empirically. Therefore the results from this study are highly interesting.

Overall, the study's methodology represents an attempt to find a middle ground between controlled stimuli presentation found in many typical neuropsychological research, and highly interactive live hyper-scanning research. This brings with it its own unique combination of strengths and weaknesses inherited from both methodologies, but can potentially help to bridge discrepant findings between both disciplines.

Given that the conclusions drawn by the authors potentially align well with ecological and enactive accounts, I think it would be interesting to **add a (very short) discussion that considers more interactional theories, rather than only focussing on sender-receiver models, such as Natural Pedagogy.** I wonder whether these theoretical accounts might align better with the authors' own conclusions, or whether the authors think that these other theories fit better with their data.

For example, if one takes the Natural Pedagogy interpretation of gaze as signalling meaningful information, one might interpret that the infants' own state (e.g. being in social learning mode) should be the best predictor of their learning, rather than aligning on a neural level. However, the way that the study is designed, it provides no feedback channel for the child to influence the adult, therefore it is not sufficiently interactive for dialogue as described by Pickering and Garrod or Participatory Sensemaking by De Jaegher and Di Paolo.

XXX Kaili & Vicky: to add brief mention of these theories/models to the introduction and discussion.

Mention these complementary theories and say that they also argue that interaction is key and that there is something ‘extra’ enacted or interacted during the conversation process which MIGHT be captured by neural synchrony.

In discussion, come back to this idea, but stress that only limited conclusions can be drawn due to uni-directional paradigm.

-------------------------------------------------------------------

COMMENT 3.2 - Gaze as a marker of structural information

One essential aspect that is not clear to me from the study's methodology is **to what extent the gaze of the video still provides structural information in the visual domain that might help infants in segmenting the visual information**. For example, there are numbers findings that have found that direct gaze or blinks do not happen at random points during speech and action:

- Direct gaze modulates speech processing and eye blinks and eyebrow movements are used to convey communicative intentions and (see studies by Hömke et al. 2017, 2025; Holler et al. 2014, Holler 2025)

- Mothers teaching children about novel object functions use gaze at event boundaries, taking into account children's knowledge with more frequent looks for younger, less knowledgeable children (e.g. Brand et al., 2007)

- Communicative signals at action boundaries can help segment actions (Kliesch, et al., 2021)

- Direct gaze and blinks can interrupt working memory (Wang & Apperly, 2016)

Taken together, gaze might represent an important source of structural information and **from the description in the methods it is not entirely clear to me what information might still be available to infants that might help their learning of auditory structures**. Whether infants use the presence or absence of communicative signals to identify learning contexts or whether they use the structural information to bootstrap the communicative function based on the structural information available within communicative signals is a major point of contention between approaches rooted in Natural Pedagogy and more ecologically rooted accounts of early infant-learning (Nomikou et al., 2016, Rączaszek-Leonardi et al., 2018; Kliesch et al., 2021; Kliesch 2025).

I might have missed this in the description of the methods, but from what I understand, the video corresponded to the actual video that corresponds to the Adult's EEG being recorded and matches the speech signal. The methods section suggested that the following manipulations were done to the sound:

> "any naturally-occurring differences in intensity were removed by digital equalisation, with loudness equalised to a playback volume of 61 dB. We further selected recordings that were closely matched for pitch, to ensure no significant differences in mean pitch across gaze conditions"

Gaze might still provide important structural information that could potentially influence infants' learning. The functions of gaze might be minimal and potentially unnoticed by the speaker or coders without a fine-grained coding. A potential control condition might consist of presenting mismatching visual and auditory information, but **I think that might be a question for a follow-up study, rather than including it in the current paper. However it would still be good to discuss which information might be available in the eye region of the stimuli, and which parts of the signal can be ruled out.**

XX Kaili: Clarify in paper that:

* Speaker did not use any intonation and the regularity of syllable delivery was controlled by a timing mechanism
* Speaker maintained direct gaze throughout (i.e. looked at the camera)
* Speaker kept head as still as possible during recording (apart from reasons of communicative cues, any movement led to light reflections on the glasses)
* No breaths during recordings – ensured consistent pacing
* Emphasise the counterbalancing of languages and gaze conditions. Particularly that gaze conditoin order was counterbalanced. 1/3 saw light first, etc.

Need to include some sort of analysis of:

* Blinks in each 20 second recording (one breath), max of 3 blinks. Report no of blinks per condition (but you can’t see dark!). Empahsise that we did control for this. Check where blinks and eyebrow movements (and any other facial movements?) occurred with respect to the word boundaries.

VL to organise additional coding of stimuli videos to provide this info

When considering the mechanisms, it would have been **nice to have more fine-grained gaze data rather than what was afforded by looking times alone**. For example, it would have been interesting to see whether infants **distributed their attention differently across the different gaze conditions and presentation orders (e.g. eyes obscured followed by visible, vs the other way around).** Alas, the paper is very complex already and this is more of a suggestion for future work.

XX Wilson to provide breakdown of learning and attention by the three orders of presentation to show that learning effect is similar across three orders.

Did Wilson add order as a variable in LMEs? Check that the effect of order was non-sign. If it was non-sig then this point is settled – we tested for an effect of order and it wasn’t there.

If it’s not non-sig then go back to Vicky for further instructions!!!!

-------------------------------------------------------------------

COMMENT 3.3 - Stimuli presentation and order

**I am a bit concerned that the visibility was manipulated in a within-subject-presentation with three different grammars**. Infants would have to ignore previously learned information, which might impede learning of the new information. **Were there any order effects in children's successful learning of the grammatical structures across multiple blocks? Maybe number of usable trials and attrition rates, and ideally a measure of successful learning for each block/running order could be added to the SI Table S6.**

XX

Wilson: Add new table after S6 to show retention rates (completed trials) per presentation order and per block. To answer the question: Does the retention rate pattern across blocks differ between presentation orders?

Wilson – add to methods clarification of the fact that in each language the words were constructed from a set of 12 consonant+vowel syllables, and these syllables were discrete for each langauge so there were no syllables that occured in more than one language.

Here are some details that might be helpful (note that these refer to the languages RECORDED, of which only a subset were USED):

Language rules:

* 3 languages (sets of ‘words’), each with versions A and B. (counterbalancing which of the test words were ‘words’ and which ‘non-words’)
* Each set constructed from one of three sets of 12 syllables (cons+vowel), using only stops, l, and r.
* Each syllable occurs in one list only, such that the three sets of materials each contain distinct sets of syllables.
* Each vowel (a e i o u) occurs 2 or 3 times in each list of syllables
* No list includes the same consonant paired with both ‘a’ and ‘u’ due to similarity of sound.
* Words were constructed from the syllable list such that no consonant or vowel occurs twice in one word.

Kaili

-------------------------------------------------------------------

COMMENT 3.4 - Reporting LMM analyses

Small point: **The LME analyses reported in the SI require some more information, e.g. see Diedrick et al., (2009) - Given the differences between packages and even versions of the same software, this should include also the software package, statistical environment and version numbers. Typically, these should also be referenced.**

XX WILSON

Xx software version

-------------------------------------------------------------------

COMMENT 3.5 - Limitations of this review

As I have not used synchrony-related analyses in my own work, I cannot comment on their appropriateness for this particular sample. I recommend soliciting additional review by someone who has published in the field.

-------------------------------------------------------------------

MINOR COMMENTS

It might be helpful for readers (and reviewers) **if it was possible to share the stimuli videos for the speaker.**

YES will share 3 x 60s videos, showing language one in each sunglasses condition.

-------------------------------------------------------------------

CONCLUSION

Overall, I think the study provides a valuable contribution to the literature. There are some potential constraints in the methodology, but they largely represent compromises that have to be made given the complexity of the data. Additionally, having a sample of mixed cultural backgrounds provides an important contribution to the literature. I think there are some interesting theoretical implications of the study's results that deserve at least a brief mention, even if an exhaustive discussion would likely go beyond the format of the journal.

My main methodological point concerns **potential structural contributions of gaze in segmenting speech information. Here, it would be very interesting to code and analyse any potential information available from the eyes (blinks, eyebrow movements) or clarify that these should be absent in the videos.** There might still be possibilities for small micro-changes influencing infants' learning and it might be helpful for readers (and reviewers) **if it was possible to share the stimuli videos for the speaker**. However, at the very least, the paper should **mention this point in the discussion section**, and the paper would need to **make it explicit which visual information might still be available in the infant.**

Finally, **more information should be provided on the attrition rates for each order/block and the statistical packages used.**

However, once these points are addressed, I think the paper would make an interesting and worthwhile contribution to the literature.

**=================================================================**

Major to do list:

1. Terminology should be changed or be consistent across paper
2. Requested discussion
3. More details/justification for current results/ parameters /statistic
4. More exploratory analyses
5. Video material / eye information open source

**=================================================================**

EDITORIAL REQUIREMENTS

The editor has requested the following actions:

1. SEX AND GENDER REPORTING

   - Include sex/gender considerations in Reporting Summary

   - Indicate if findings apply to only one sex/gender in title/abstract

   - Report whether sex/gender was considered in study design

   - Provide disaggregated data in source files

   Related manuscript sections:

   [To be filled in]

2. DATA AND CODE AVAILABILITY

   - Include Data Availability section after Methods

   - Include Code Availability section after Data Availability

   - Ensure GitHub code link is working

   - Consider using figshare repository for data

   Related manuscript sections:

   [To be filled in]

3. FIGURES - Replace bar graphs with distribution plots

   - Show all data points for n<10

   - Use box-and-whisker or violin plots for larger samples

   - Include measures of centrality, dispersion, error bars

   Affected figures:

   [To be filled in]

4. ORCID

   - All corresponding authors must link ORCID IDs

   - Notify all co-authors to link ORCIDs prior to acceptance

   Action required:

   [To be filled in]

5. STATISTICAL REPORTING

   - Follow Nature Communications statistical guidance

   - Consider sensitivity analyses per Lakens (2022)

   Related manuscript sections:

   [To be filled in]

**===================================================================**

End of Document

Reference temporally store here

Benitez, V. L., Bulgarelli, F., Byers‐Heinlein, K., Saffran, J. R., & Weiss, D. J. (2020). Statistical learning of multiple speech streams: A challenge for monolingual infants. *Developmental Science*, *23*(2), e12896. https://doi.org/10.1111/desc.12896

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). *Statistical treatment of looking-time data*. *52*(4), 521.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Hay, J. F., & Saffran, J. R. (2012). Rhythmic Grouping Biases Constrain Infant Statistical Learning. *Infancy*, *17*(6), 610–641. https://doi.org/10.1111/j.1532-7078.2011.00110.x

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8‐Month‐Old Infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

Thiessen, E. D., & Saffran, J. R. (2007). *Learning to Learn: Infants’ Acquisition of Stress-Based Strategies for Word Segmentation*.