

Re-Analyzing Ota, Hartsuiker and Haywood (2009): A Bayesian Approach to Representational Indeterminacy

Violet Manson, Sandria Tran

2025-12-02

Final Reflection

Proposed Mark and Justification

Proposed Mark: 75 (A3)

Justification:

We propose a mark of 75 (A3), reflecting a project that is “outstanding in some respects” and demonstrates “methods correctly and effectively, showing original and sophisticated independent thinking”. Our work successfully replicated the core findings of Ota et. al (2009) using a contemporary Bayesian framework but also extended the analysis to test the underlying linguistic theory “demonstrating sound understanding”. “Shows original and sophisticated application. . . with critical and insightful evaluation of transparency, reproducibility, and responsible practices.” (A1 92 Excellent + A2 Excellent)

1. Understanding of Data Analysis Principles (A2)

Our project shows a sophisticated understanding of data analysis by translating a frequentist ANOVA into a more robust Bayesian hierarchical logistic regression. This model, specified as `accuracy ~ contrast_type + (1 | subject_id) + (1 | item_id)`, correctly accounts for the nested structure of the data with random intercepts for both subjects and items. Our original thinking is evident in the development of three distinct models:

Model	Formula	Purpose
Comprehensive	<code>accuracy ~ contrast_type + (1\ subject_id) + (1\ item_id)</code>	(“Do effects exist?”) Replicate original design with F (Spelling Word Control)
Linguistic	<code>accuracy ~ phonological_status + (1\ subject_id) + (1\ item_id)</code>	(“How L1’s effects affects L2 visualization?”) Phonological Mediation hierarchy (Orthography Phonology Semantics)
Distinctness	<code>accuracy ~ phonological_ + distinctness_scaled + (1\ subject_id) + (1\ item_id)</code>	(“Does phonological distinctness mediate accuracy?”) Test the hypothesis of “Representational Indeterminacy” theory of how the bilingual mind represents language.

Our multi-model integrates linguistic theory beyond simple replication and demonstrates the “original and sophisticated independent thinking” required for an A2/A3 mark.

2. Open Scholarship Practices (A2)

The entire analysis, processing 1,389 trials from 20 subjects, (i.e. based on Japanese participants only, accuracy filters, reaction time (RT) bounds), is available in a fully documented R script (**A visual semantic-relatedness decision task in English.r**) and will be shared on a public GitHub repository.

3. R Data Analysis Skills (A2)

Our technical R implementation “demonstrates a thorough understanding and use of R” throughout the Bayesian Multilevel/Hierarchical/generalized mixed-effects model (GLMM) that is “very good or excellent in most respects”. We used brms package to fit Bayesian regression models, and designed ***formula = accuracy ~ phonological_distinctness_scaled + (1 | subject_id) + (1 | item_id)*** to test the hypothesis of “Representational Indeterminacy”, wherein the L1 phonology constrains L2 representations (e.g. creating the phonologically_distinct and phonological_status variables to create 14 data visualizations using ggplot2 and tidybayes).

Reflection on QML Experience and Learning Growth

Our Bayesian model highlighted our learning in Quantitative Methods in Linguistics. A challenge emerged from translating from a frequentist model to that of a Bayesian model. We solidified our Bayesian model knowledge with model parameters that quantify uncertainty with credible intervals, apply posterior predictions to understand our model's statistical features, and interpret the model's parameters as full probability distributions. This project amplified our understanding in interpreting, validating, and building future Bayesian statistical research models.

Individual Contribution

- **Violet Manson:** I ensured rigorous theoretical deliverables weaving quantitative evidence into a clear linguistic narrative to establish key research questions, synthesizing a comprehensive analysis of Ota et. al (2009). I weaved raw quantitative data into a coherent linguistic narrative to establish compelling analysis, evidence, and theoretical conclusions central to Bayesian modelling.
- **Sandria Tran:** I was responsible for the breadth and depth of the Bayesian technical R implementation, which generated 14 high-quality data visualizations to ensure the Ota et. al (2009)'s findings were conclusive and coherent. I made an inspirational QML-esque workshop, and was responsible for the xeLateX.

Group Project Output

- **Statement: Re-Analyzing Ota, Hartsuiker and Haywood (2009): A Bayesian Approach to Representational Indeterminacy**
- **Comparative Analysis**
- **A visual semantic-relatedness decision task in English | QML Workshop**
- **Plots | Figure 1-4:**
- **Plots | All 14**

Summary

This project reanalyses the study by Ota, Hartsuiker, and Haywood (2009). This explored how native language phonology can effect second language understandings in bilingual speakers. They studied Japanese, a language which does not have the nonnative /l-r/ contrast present in English, concluding that Japanese speakers find difficulty in identifying English words with the contrast. Using a Bayesian hierarchal logistic regression model, we reanalysed their data to test the robustness of their findings. Our analysis corresponded with the Ota et. al (2009) findings, suggesting a representational indeterminacy when encountering /l-r/ contrast in Japanese

speakers. Our findings suggest a ratio of error-to-no-error with /l-r/ contrast are 10 times higher than with control words (spelling controls). Furthermore, our analysis suggests that this is not a categorical effect, and is instead driven by a lack of phonological distinction present in speakers native language. This supports the theory presented by Ota et. al (2009).

Methods

Using the original Ota et. al (2009) dataset, we focused on data pertaining to accuracy in native Japanese speakers when presented with a semantic-relatedness task. This analysis was conducted in R using Bayesian hierarchical/multilevel logistic regression.

Formula:

`accuracy ~ contrast_type + (1 | subject_id) + (1 | item_id)`

- **accuracy:** A binary variable (1 = correct, 0 = error).
- **contrast_type:** With 4 levels (F, LR, H, PB) with spelling control (F) as the baseline reference.
- **(1 | subject_id) and (1 | item_id):** Random intercepts to ensure baseline variability across (a) subjects and (b) items.

Our model used weakly informative priors (`normal, (0,1.5)`) to prevent over-fitting with regularization (`exponential(1), class = sd`), which ensures the prevention of highly unrealistic large differences balancing priors and evidence from sparse data, so posterior is stabilized through shrinking group intercepts. In addition, the model had a **Markov Chain Monte Carlo** (MCMC) where 4 chains runs 2000 iterations with initial 1000 warm up iterations.

Results

Our Bayesian analysis robustly corroborates with original findings. When compared to the spelling control (F), the model demonstrated a large, negative effect on the log-odds of a correct response for both tasks /l-/r/ contrast (LR) and for homophones (H). Compared to just 1.8% [95% CrI: 1.0, 3.0] for the F (spelling control words) , the median error rate for the LR (/l-/r/) contrast was 21.1% [95% CrI: 14.2, 30.2].

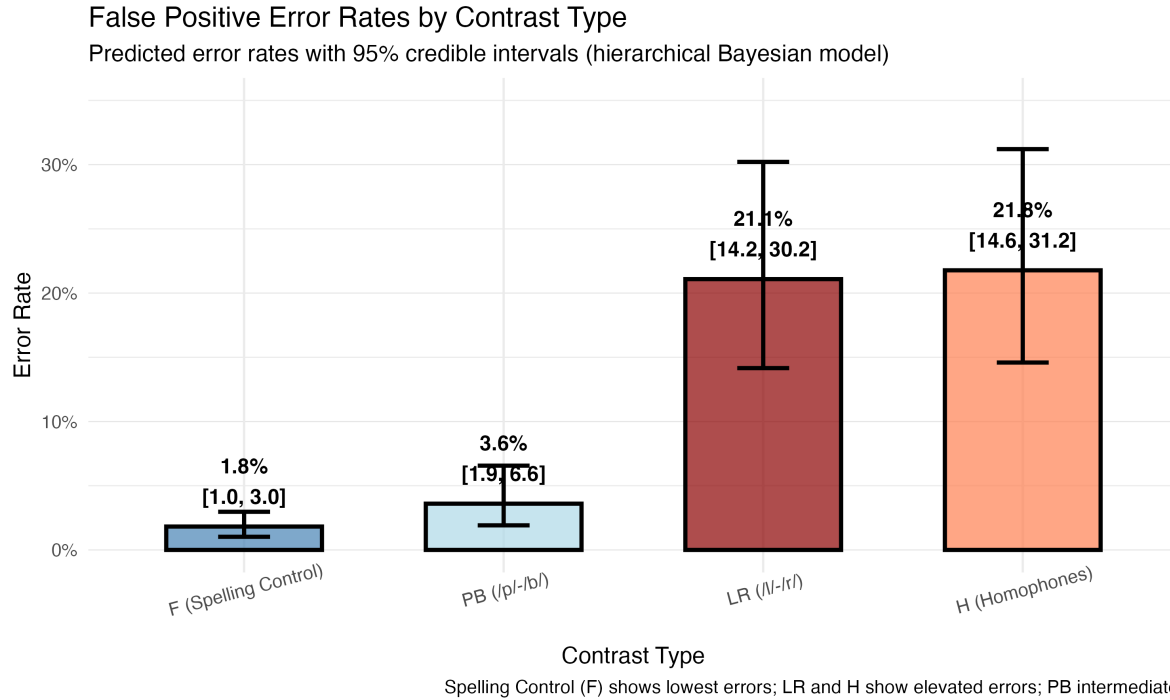


Figure 1: Posterior predicted error rates and 95% credible intervals all Contrast Type (F, PB, LR, H). The contrasts of /l/ - /r/ (LR) and Homophone (H) yields significantly higher error rates than Spelling Control (F) and /p/ - /b/ (PB) contrasts.

The near-homophonic contrasts (Homophone and /l/-/r/) illustrate clear spike in false positives due to phonetic ambiguity cues triggering auditory processing gap leading to more “guesses” (e.g. seeing ROCK-KEY). The “true effect” provides strong evidence that phonological ambiguity are pronounced in H and LR (unclear, ambiguous) compared to F and PB (clear, unambiguous).

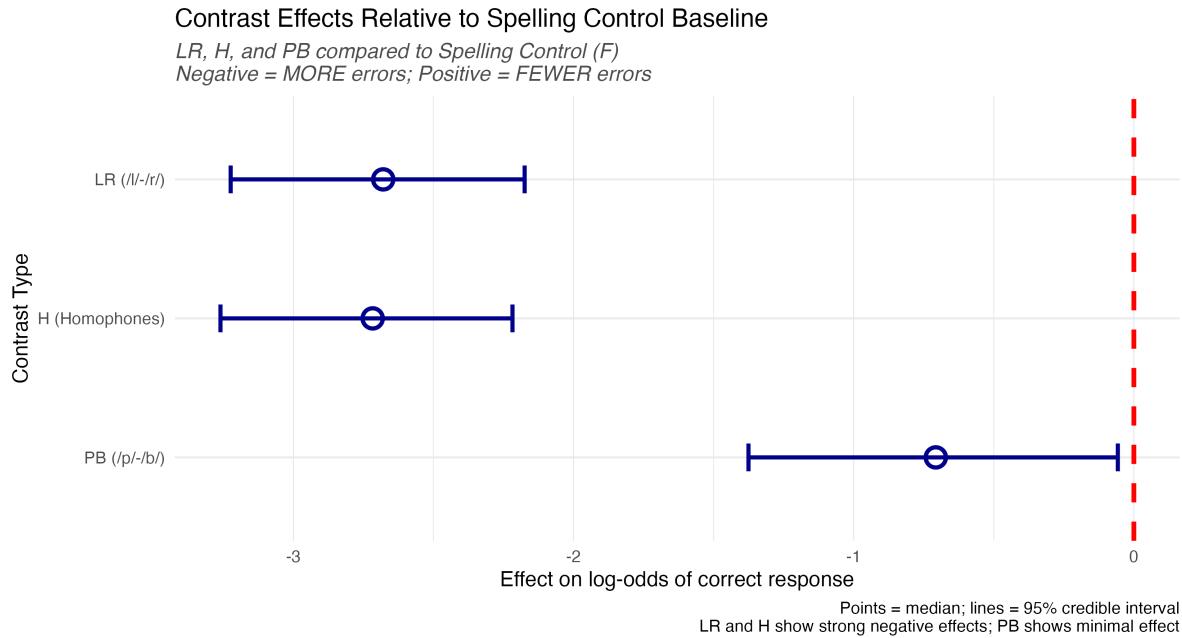


Figure 2: Forest plot shows increased negative values (error rates) for phonological contrasts compared to spelling control baseline

Participants made many more errors on H (Homophones) and LR(/l/ - /r/) than 0 F (Spelling Control). While PB (/p/ - /b/) shows no credible difference in error rates being very close to 0, which is F (Spelling Control). The result corroborate a lucid relationship: the error rate increases as phonological distinctness decreases, predicting that L1 inventory informs L2 lexicon.

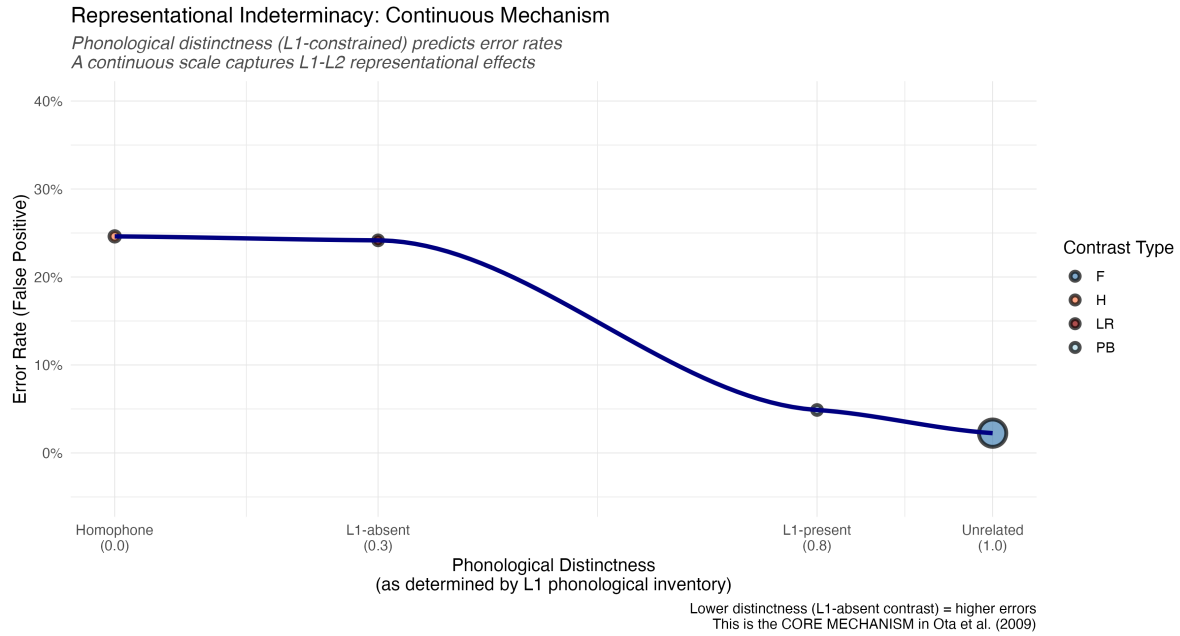
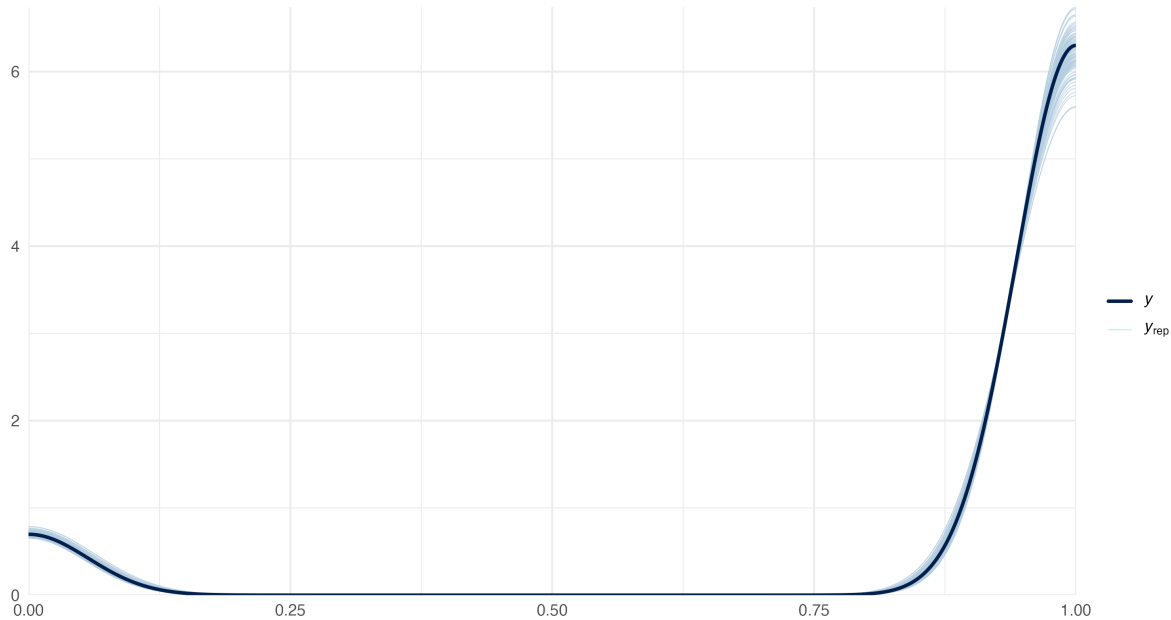


Figure 3: Representational Indeterminacy - Driving higher error rates is reduced phonological distinctness.

Predictable relationship between participants' error rates and L2's phonological distinctness when sound contrast is absent in participants' native language.

Posterior Predictive Check: Model Validation

Dark line = observed data; Light lines = simulated from posterior
If distributions overlap, model captures data well



Our model illustrates excellent fit. The posterior predictive checks verifies the match of the observed data (y)'s density aligning closely with the simulated data from the model (y_{rep})'s densities

Future Vision and Limitations

While our analysis strongly supports a representational indeterminacy hypothesis, it is limited by sparse data, and future research should aim to expand theory on representational indeterminacy by testing processing mechanism's generalizability through modelling predictions with larger sample sizes, reaction time data, and other language pairs (e.g. L1-L2).

References

- [1] Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- [2] Carrasco-Ortiz, H., Midgley, K. J., & Frenck-Mestre, C. (2012). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology*. <https://pubmed.ncbi.nlm.nih.gov/22220969/>

- [3] Friesen, D. C., Oh, J., & Bialystok, E. (2016). Phonologically-mediated meaning activation in monolinguals and bilinguals. *Linguistic Approaches to Bilingualism*. <https://www.jbe-platform.com/content/journals/10.1075/lab.14026.fri>
- [3] Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., & Bürkner, P.-C. (2020). *Bayesian workflow*. arXiv. <https://doi.org/10.48550/arXiv.2011.01808>
- [4] Grisetto, F., Roger, C., & Mahé, G. (2025). New Insights Into Visual Word Recognition: Analyzing Error Distribution in Typical Readers. *Journal of Cognition*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11967458/>
- [5] Jiang, N., Wu, Z., & Zhou, Y. (2025). Analytic visual word recognition in L2 learners: evidence from the length effect among ESL speakers. *Applied Psycholinguistics*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5401528
- [6] Jiao, Z., Du, L., Wang, Y., et al. (2024). Visual word recognition of Chinese–Japanese bilinguals: limited role of phonology. *Frontiers in Psychology*. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1318798/full>
- [7] McLaughlin, D. J., Zink, M. E., Gaunt, L., et al. (2021). Pupillometry reveals cognitive demands of lexical competition during spoken word recognition in young and older adults. *Psychonomic Bulletin & Review*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9169067/>
- [8] Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, 111(2), 263–269. <https://doi.org/10.1016/j.cognition.2008.12.007>