EEG Many Pipelines - The Code Mechanics

Sebastian Speer[1], Antonio Schettino[2,3], & Ana Martinovici[4]

[1] Social Brain Lab, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands

[2] Erasmus Research Services, Erasmus University Rotterdam, Rotterdam, The Netherlands

[3] Institute for Globally Distributed Open Research and Education (IGDORE), Sweden

[4] Rotterdam School or Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Author Note

Authorship order was randomly determined via the `sample` function in *R*.

AS preprocessed the data and performed ERP analysis (research questions 1, 2a, 3a, 4a). **SS** preprocessed the data and performed time-frequency analysis (research questions 2b, 2c, 3b, 4b). **AM** was responsible for project management, GitHub repository, and reproducibility. **AS**, **SS**, and **AM** wrote the report.

Correspondence concerning this article should be addressed to Ana Martinovici, Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands. E-mail: martinovici@rsm.nl

EEG Many Pipelines - The Code Mechanics

1 Introduction

This report describes the methods and results of our contribution to the **EEGManyPipelines** project.

We registered our team as **The Code Mechanics**[1] on September 15th 2021 and received a confirmation email five days later. Instructions on how to download the EEG data and which research hypotheses to test were sent on October 8th 2021. A copy can be found at `data_in_repo/original_data/instructions`.

We received the link to the submission portal on May 2nd 2022, submitted all relevant materials and filled out three questionnaires before the deadline (May 15th 2022).

2 Methods

2.1 Preprocessing

2.1.1 ERP. For each participant, the continuous EEG data was assigned electrode coordinates and filtered with two consecutive one-pass, zero-phase, non-causal FIR filters (Hamming window with 0.0194 passband ripple and 53 dB stopband attenuation): (*i*) high-pass: 0.1 Hz (-6 dB cutoff frequency: 0.05 Hz), length 16897 samples; (*ii*) low-pass: 40 Hz, upper transition bandwidth 10 Hz (-6 dB cutoff frequency: 45.00 Hz), length 169 samples (Widmann, Schröger, & Maess, 2015). Bad or noisy channels were detected using several approaches implemented in the *PREP* pipeline (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015), including: (*i*) flat or missing signal; (*ii*) "bad-by-high-frequency-noise" (frequency components above 50 Hz considerably higher than median channel noisiness, as calculated using a robust Z-scoring method and default threshold $z > 5$); (*iii*) "bad-by-correlation" (maximum correlation with another

[1] A homage to **The Organic Mechanic**, a doctor in the movie **Mad Max: Fury Road**.

channel below the default $r = .4$, fraction of bad correlation windows above the default threshold of 0.01); (*iv*) "bad-by-deviation" (amplitude deviates from the median channel amplitude, as calculated using a robust Z-scoring method and default threshold $z > 5$); and (*v*) signal prediction based on signals and spatial locations of other channels lower than default threshold of $r = .75$ (random sample consensus approach, *RANSAC*; Fischler and Bolles (1987)).[2] Channels identified as noisy were removed from the data and subsequently interpolated via a spherical spline procedure (Perrin, Pernier, Bertrand, & Echallier, 1989).

Afterwards, the EEG data were re-referenced to the average signal across channels. Ocular artifacts were then corrected by means of independent component analysis via the `picard` algorithm (Ablin, Cardoso, & Gramfort, 2018) and the resulting components were correlated with activity recorded from the EOG channels, in order to identify which components best represented horizontal eye movements and blinks. The components that correlated the highest with the EOG channels were then removed from the EEG data before re-conversion into channel space.

The EEG data was subsampled by a factor of four (i.e., from 512 Hz to 128 Hz), segmented into epochs extending from -200ms to +500ms time-locked to scene onset, and baseline corrected using the pre-stimulus interval. The epoched data was then subjected to *Autoreject*, an automated artifact detection algorithm based on machine-learning classifiers and cross-validation to estimate the optimal peak-to-peak threshold (Jas, Engemann, Bekhti, Raimondo, & Gramfort, 2017) [3]. This algorithm was implemented to remove artifacts not identified in previous preprocessing steps.

2.1.2 TFR. Signal preprocessing for time-frequency analysis followed a similar procedure, with the following exceptions: (*i*) high-pass FIR filter: 1 Hz (-6 dB cutoff frequency: 0.50 Hz), length 1691 samples; (*ii*) continuous data segmented into 800 ms epochs (-300ms to +500ms time-locked to scene onset) and baseline corrected using the

[2] For more details, see the documentation of `NoisyChannels`.

[3] For more details, see the documentation of `autoreject`.

pre-stimulus interval.

The preprocessed data were then submitted to a Morlet wavelet analysis to transform the data into the time-frequency domain with 18 log-scaled frequency bins ranging from 4 to 40 Hz, to increase sensitivity at lower frequency ranges. To optimize both spectral and temporal resolution, the number of cycles to include in the sliding time window were defined by dividing each individual frequency by two. After transforming the data to the time-frequency domain, the data were decimated by a factor of two (sampling every second time point) to increase computational efficiency.

2.1.3 Epoching. For both ERP and TFR analysis, the clean epoched data were separated into 8 conditions:

- **RQ1**

    1. *manmade*: scenes categorized as `man-made`, presented for the first time or previously (`new` and `old`), subsequently remembered or forgotten (`subsequent_remembered` and `subsequent_forgotten`), excluding `NA` in behavioral responses (although scene category is independent from response, `NA` may reflect drops in attention and, consequently, incomplete stimulus perception)
    2. *natural*: scenes categorized as `natural`, `new` and `old`, `subsequent_remembered` and `subsequent_forgotten`, excluding `NA` in behavior

- **RQ2**

    3. *new*: `man-made` and `natural` scenes, presented for the first time (`new`), `subsequent_remembered` and `subsequent_forgotten`, excluding `NA` in behavior
    4. *old*: `man-made` and `natural` scenes, presented previously (`old`), `subsequent_remembered` and `subsequent_forgotten`, excluding `NA` in behavior

- **RQ3**

5. *old-hit*: `man-made` and `natural` scenes, presented previously (`old`), successfully recognized as such (`hit`), can include `NA` in memory (the image has been successfully categorized as old, regardless of whether it is recognized as such in subsequent presentations)
6. *old-miss*: `man-made` and `natural` scenes, presented previously (`old`), not recognized as such (`miss`), can include `NA` in memory

- **RQ4**

7. *remembered*: `man-made` and `natural` scenes, `new` and `old`, `subsequent_remembered`, include all behavior
8. *forgotten*: `man-made` and `natural` scenes, `new` and `old`, `subsequent_forgotten`, include all behavior

3 Analysis

3.1 RQ1

The *EEGManyPipelines* team asked us to test the following:

RQ1. There is an effect of scene category (i.e., a difference between images showing man-made vs. natural environments) on the amplitude of the N1 component, i.e. the first major negative EEG voltage deflection.

To address this question, we identified the N1 ERP component by visually inspecting topographies and waveforms of the grand-average signal (i.e., collapsed across all participants and both *man-made* and *natural* conditions). This unbiased *collapsed localizer* approach (Luck & Gaspelin, 2017, p. 150) revealed a negative deflection at a region of interest (*ROI*) comprising electrodes *PO7*, *PO3*, *O1*, *PO4*, *PO8*, and *O2* (see Figure 1), and a time window between 130 - 180 ms after stimulus onset (see Figure 2).

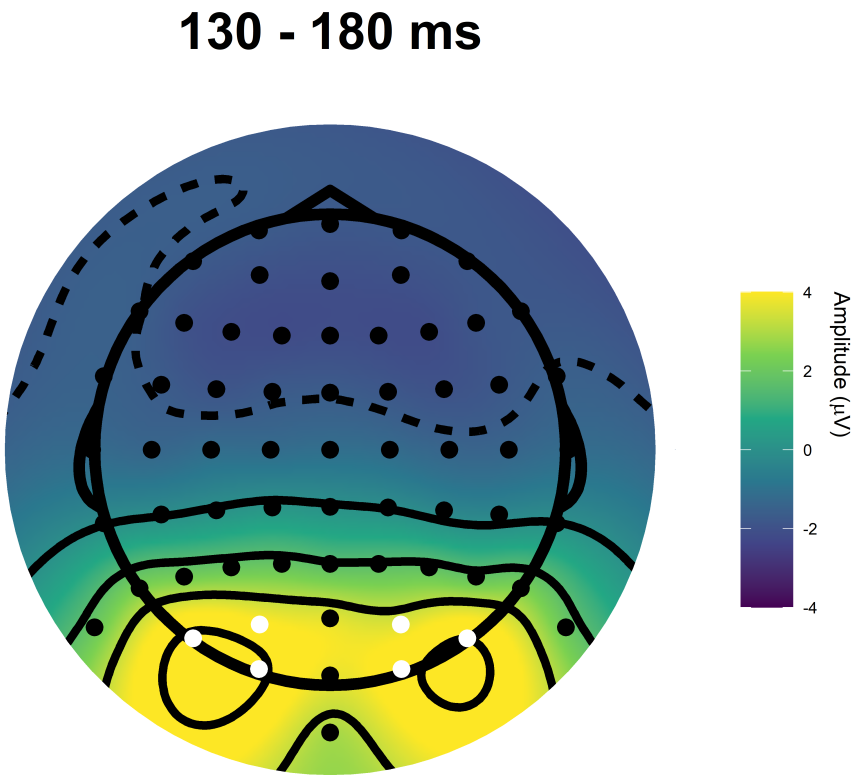

Figure 1. Collapsed localizer (i.e., amplitude averaged across participants and conditions between 130 - 180 ms after stimulus onset) used to identify the electrodes best representing the expected topography of the N1 ERP component (*PO7*, *PO3*, *O1*, *PO4*, *PO8*, *O2*).

Participant- and condition-specific N1 amplitude was extracted by averaging values recorded at this ROI and time window (see Figure 3).

Subsequently, we fit a Bayesian multilevel linear model on N1 amplitude values, with *condition* (2 levels: *man-made*, *natural*) as *constant* (a.k.a. fixed) effect and participant and trial as *varying* (a.k.a. random) effects. We allowed intercepts and slopes to vary as a function of participant and trials, to model general and condition-specific inter-individual differences. As likelihood function, we chose a Gaussian distribution.

An important aspect of Bayesian analysis is the choice of priors (e.g., Natarajan & Kass, 2000). Given the susceptibility of the electrophysiological signal to inter-individual differences (e.g., skull thickness, skin conductance, or hair), we decided to base our priors

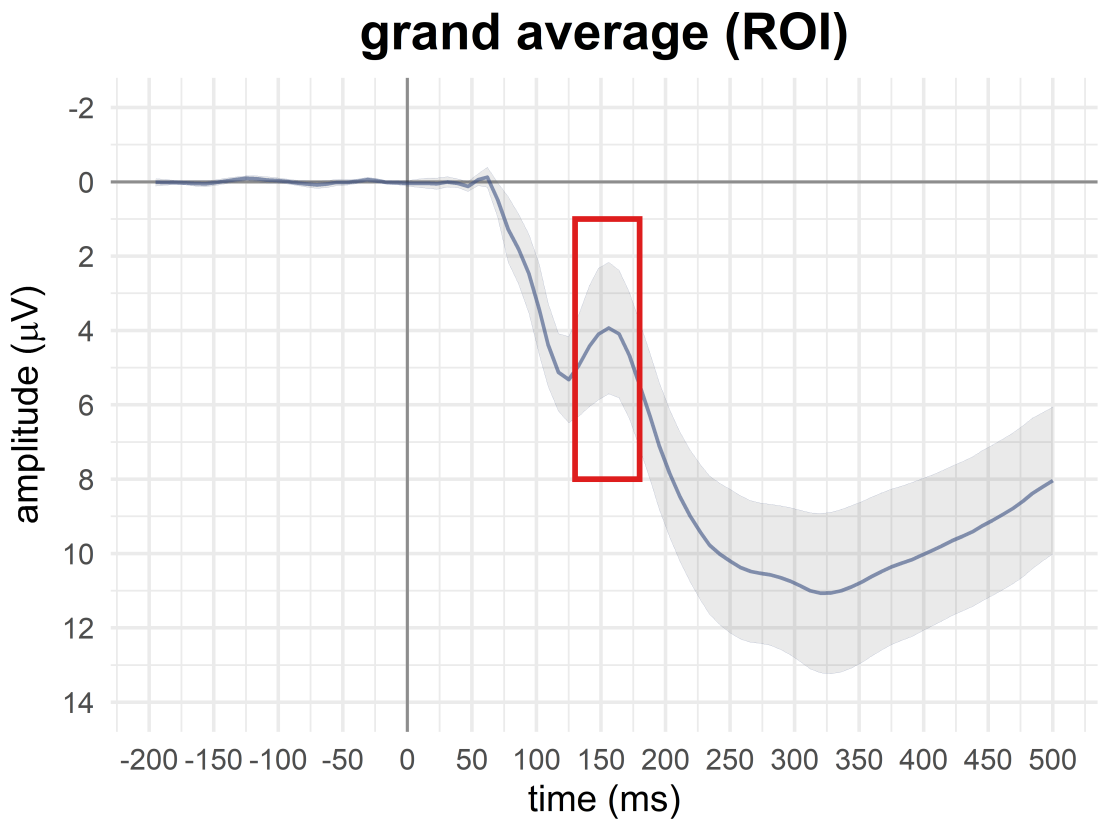

Figure 2. Collapsed localizer (i.e., amplitude averaged across participants and conditions at electrodes *PO7*, *PO3*, *O1*, *PO4*, *PO8*, *O2*) used to identify the time window of the N1 ERP component between 130 - 180 ms after stimulus onset (red box).

on the current data by visually inspecting the *collapsed localizer* (see above). For the main analysis, we placed **informative priors** on the *intercept* – a normal distribution with mean $\mu = 4$ and standard deviation $\sigma = 2$: $Normal(4, 2)$ – and β coefficient, $Normal(0, 1)$. To assess whether our chosen informative prior would bias parameter estimates (and, consequently, the interpretation of the results; Depaoli and van de Schoot (2017)), we ran the same multilevel linear model with **weakly informative** priors (*intercept*: $Normal(4, 4)$; β: $Normal(0, 4)$) and **uninformative** priors (*intercept*: $Normal(4, 10)$; β: $Normal(0, 10)$). We anticipate that the choice of prior had negligible effects on the posterior distributions, because the influence of the prior washes out with a large amount of data (Edwards, Lindman, & Savage, 1963).

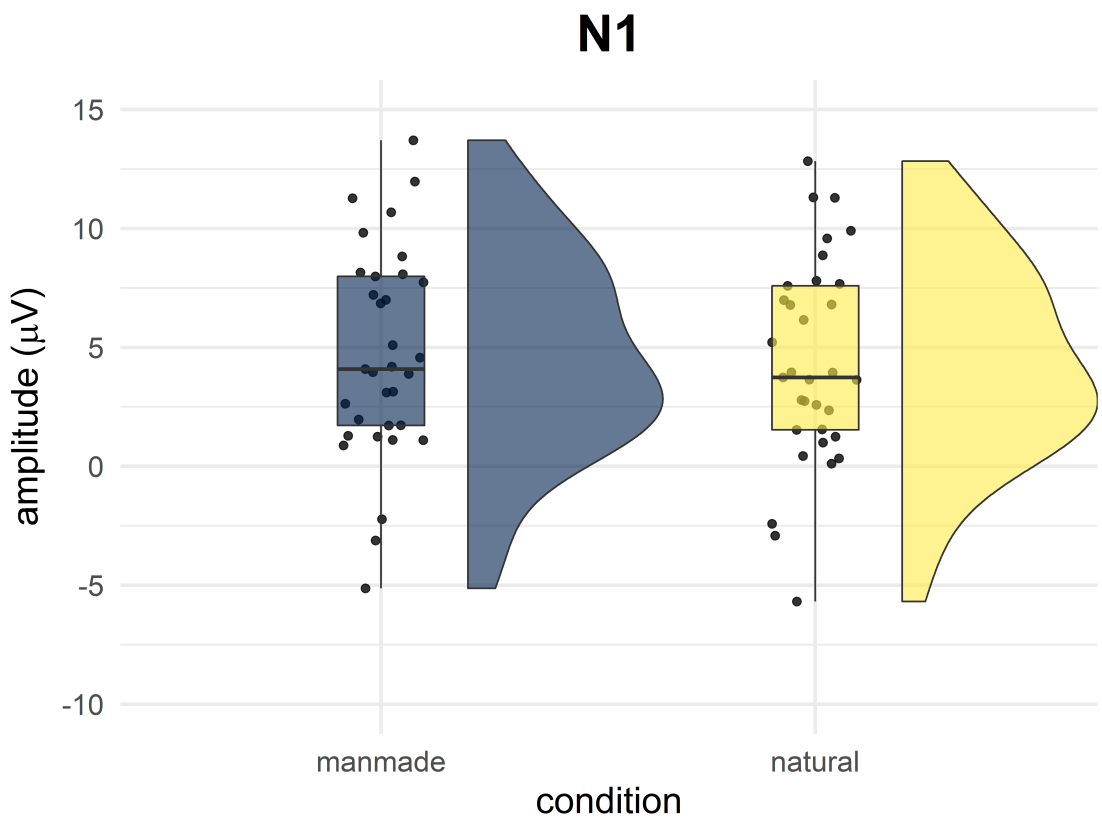

Figure 3. Raincloud plot (Allen et al., 2021) showing trial-averaged N1 amplitude values – extracted from the electrode ROI and time window identified via visual inspection of the collapsed localizer –, separately for each participant and condition. Please note that trial-averaged data are shown here only for illustration purposes, whereas statistical analyses were performed on trial data.

Since we had no prior knowledge regarding the standard deviation of participant and trials, we placed a **weakly informative prior** on these varying effects: a t-distribution with degrees of freedom $\nu = 3$, location $\mu = 0$, and scale $\sigma = 2$, $Student(3, 0, 2)$.

Models were fitted in *R* using the `brms` package (Bürkner, 2018), which employs the probabilistic programming language *Stan* (Carpenter et al., 2017) to implement a Markov chain Monte Carlo (MCMC) algorithm (i.e., No-U-Turn sampler; Homan and Gelman (2014)) to estimate posterior distributions of the parameters of interest. Four MCMC chains with 4000 iterations (2000 warm-up) and no thinning were run to estimate

parameters in each of the fitted models. Model convergence was assessed as follows: (*i*) visual inspection of trace plots, rank plots, and graphical posterior predictive checks (Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019); (*ii*) Gelman-Rubin $\hat{R}$ statistic (Gelman et al., 2013) – comparing the between-chains variability to the within-chain variability – between 1 and 1.05 (see also Nalborczyk, Batailler, Lœvenbruck, Vilain, & Bürkner, 2019).

Posterior distributions of the model parameters were summarized using the mean and 95% credible interval (CI). Differences between conditions were calculated by computing the difference between posterior distributions of the respective conditions.

Statistical inference was performed using the **HDI + ROPE** decision rule (Kruschke, 2018): posterior differences were accepted or rejected against a null hypothesis considering a small effect as "practically equivalent to zero" (Region of Practical Equivalence; *ROPE*). To mitigate the inevitable subjectivity intrinsic in arbitrarily choosing the values "practically equivalent to zero", we employed a range of plausible ROPEs, from $\pm 0.05\ \mu V$ to $\pm 0.5\ \mu V$ in steps of $0.01\ \mu V$. If the percentage of the posterior differences within the full ROPE was smaller than 5%, the null hypothesis was rejected (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019).

3.2 RQ2a

The *EEGManyPipelines* team asked us to test the following:

> *RQ2a. There are effects of image novelty (i.e., between images shown for the first time/new vs. repeated/old images) within the time-range from 300–500 ms on EEG voltage at fronto-central channels.*

To address this question, for each participant and *new*/*old* condition we averaged activity in the time window between 300 - 500 ms after stimulus onset and electrode ROI

FC1, *FC2*, *FCz*, i.e., electrodes strictly classified as fronto-central according to the international 10/10 electrode system (Chatrian, Lettich, & Nelson, 1985).

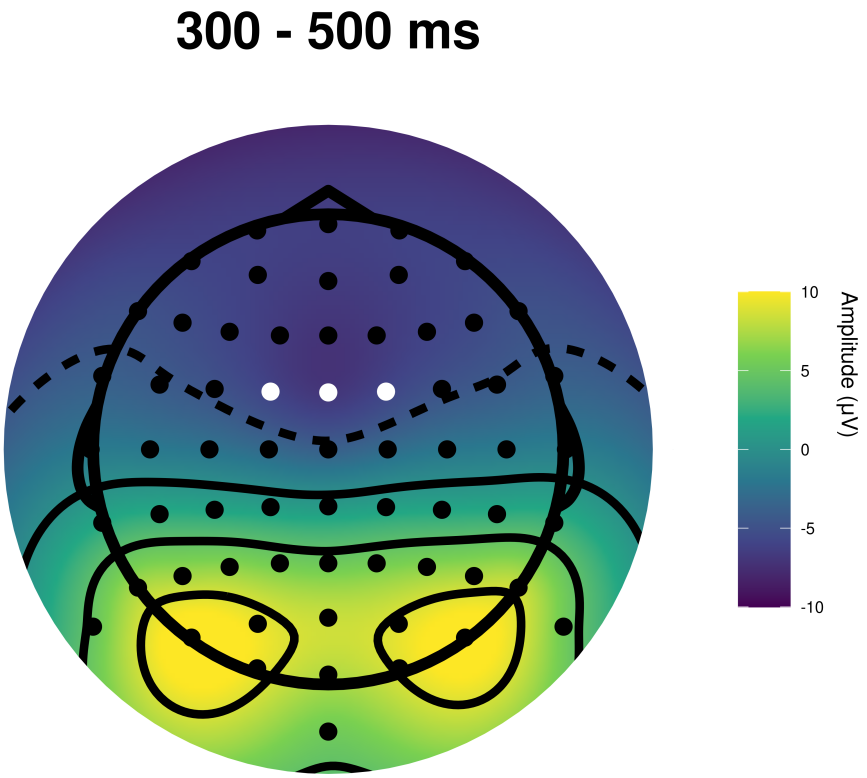

Figure 4. Collapsed localizer (i.e., amplitude averaged across participants and conditions between 300 - 500 ms after stimulus onset) used to confirm that the selected electrode ROI (*FC1*, *FC2*, *FCz*) accurately reflects the expected negative deflection.

Amplitude values recorded at this ROI and time window were averaged across trials, separately for each participant and condition (see Figure 4).

We fit Bayesian multilevel linear models following the procedure described in section 3.1, with the following differences in priors: (*i*) **informative priors**: $Normal(-8, 2)$ on intercept, $Normal(0, 1)$ on β; (*ii*) **weakly informative priors**: $Normal(-8, 4)$ on intercept, $Normal(0, 4)$ on β; (*iii*) **uninformative priors**: $Normal(-8, 10)$ on intercept, $Normal(0, 10)$ on β.

Summaries of the posterior distributions of model parameters and statistical inference

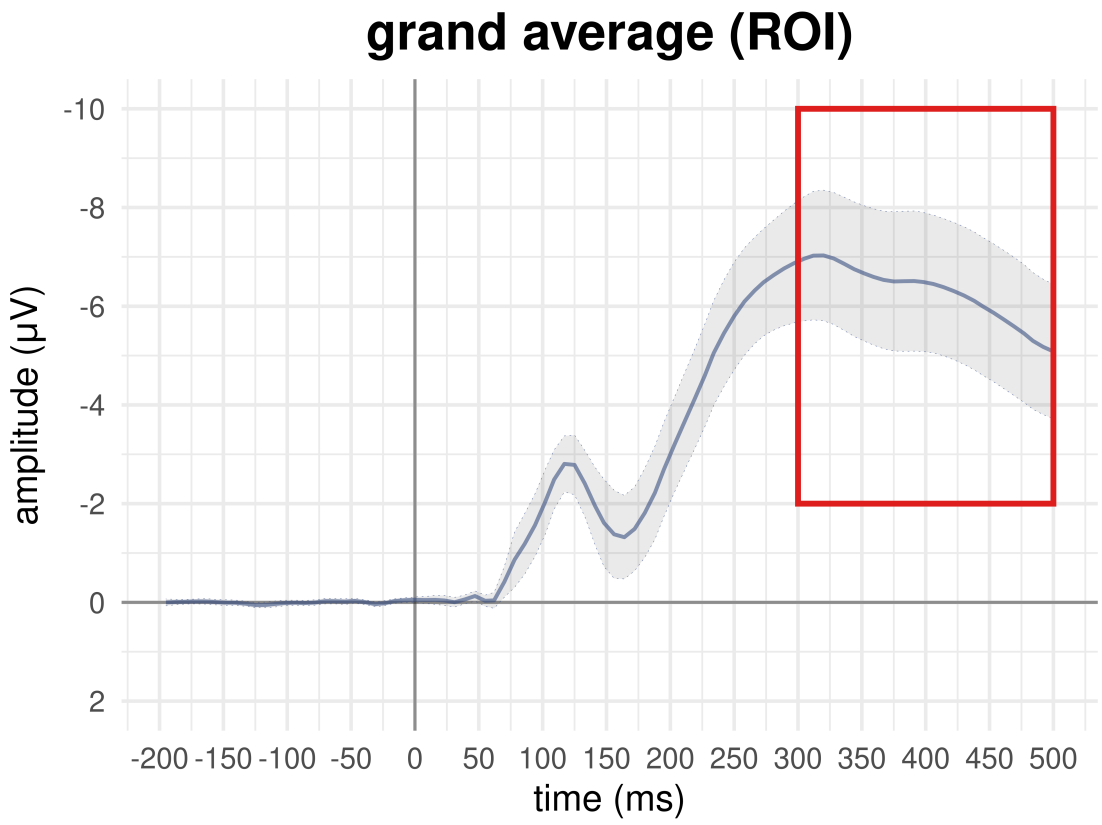

Figure 5. Collapsed localizer (i.e., amplitude averaged across participants and conditions at electrodes *FC1*, *FC2*, *FCz*) used to identify the time window of the expected negative deflection between 300 - 500 ms after stimulus onset (red box).

were performed as described in section 3.1.

3.3 RQ2b and RQ2c

The *EEGManyPipelines* team asked us to test the following:

RQ2b. There are effects of image novelty (i.e., between images shown for the first time/new vs. repeated/old images) within the time-range from 300–500 ms on theta power at fronto-central channels.

RQ2c. There are effects of image novelty (i.e., between images shown for the

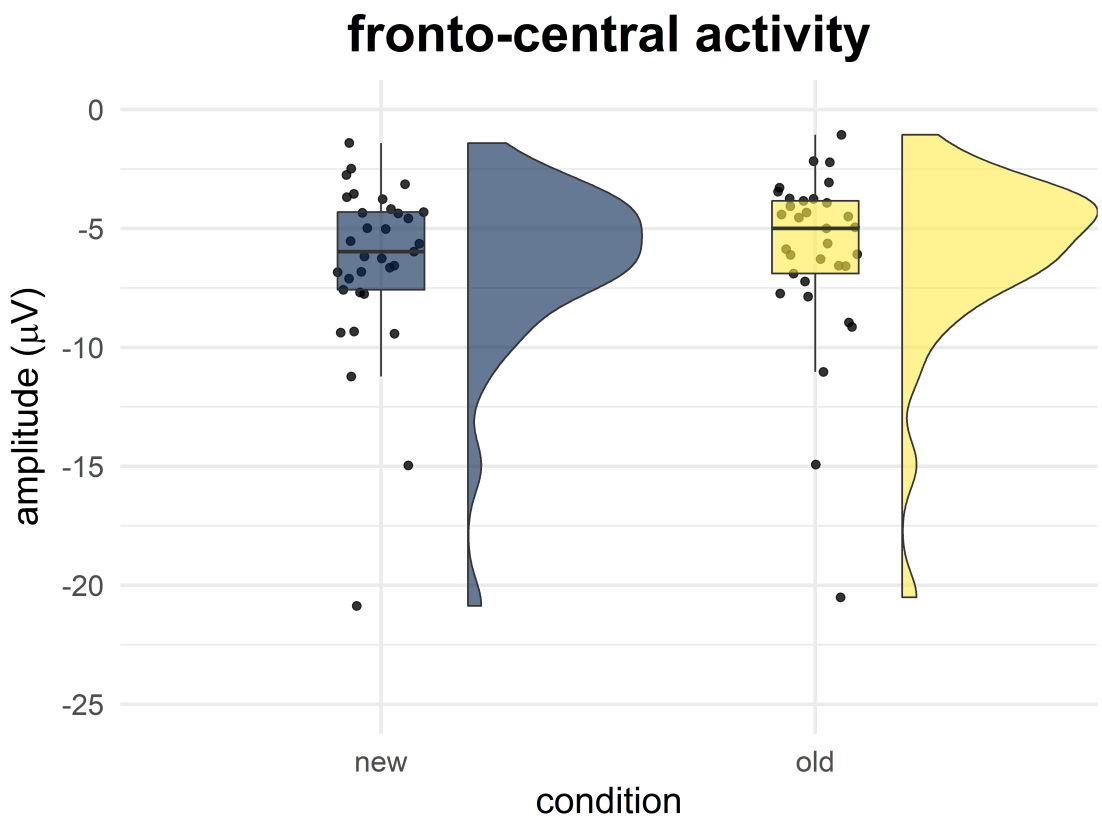

Figure 6. Raincloud plot showing trial-averaged amplitude values – extracted from the specified electrode ROI and time window –, separately for each participant and condition. Please note that trial-averaged data are shown here only for illustration purposes, whereas statistical analyses were performed on trial data.

> *first time/new vs. repeated/old images) within the time-range from 300–500 ms on alpha power at posterior channels.*

To address these questions, we conducted a multilevel analysis contrasting the EEG data from trials with *old* images against trials with *new* images. At the first level (i.e., the participant level), we computed the averaged time-frequency maps for each of the two conditions. We then tested the resulting averaged maps at the second level for significant group effects using a paired-sample *t*-test. We used cluster-based permutation testing, due to its stringent control of familywise error rates (Maris & Oostenveld, 2007). Specifically,

for every sample across the three channels, we quantified the experimental effect by a t-value. Selection of samples for inclusion in a cluster was implemented using threshold-free cluster enhancement (*TFCE*) (Smith & Nichols, 2009). TFCE eliminates the free parameter initial threshold value that determines which points are included in clustering by approximating a continuous integration across possible threshold values with a standard Riemann sum. We subsequently clustered selected samples in connected sets based on temporal and spectral adjacency and computed cluster-level statistics by taking the sum of the t-values within every cluster. Subsequently, we performed permutation testing using the Monte Carlo method (1000 permutations) to compute the significance probability of our observed effect (Maris & Oostenveld, 2007) ($\alpha = 0.05$). This analysis results in a cluster of adjacent data points across time, frequencies, and channels, which significantly differs in activity between old and new images.

To test for differences in **theta** (θ) power at fronto-central channels, we focused on the frequency range from 4 - 8 Hz and all frontocentral channels (FC1, FCz, FC2).

To test for differences in **alpha** (α) power at posterior channels, we focused on the frequency range from 8 - 13 Hz and all posterior channels (P7, P5, P3, P1, P2, P4, P6).

3.4 RQ3a

The *EEGManyPipelines* team asked us to test the following:

> *RQ3a. There are effects of successful recognition of old images (i.e., a difference between old images correctly recognized as old [hits] vs. old images incorrectly judged as new [misses]) on EEG voltage at any channels, at any time.*

We followed the same analysis approach described in section 3.3 but on time-series data, comparing *old-hit*/*old-miss* conditions and including all timepoints and channels.

3.5 RQ3b

The *EEGManyPipelines* team asked us to test the following:

> *RQ3b. There are effects of successful recognition of old images (i.e., a difference between old images correctly recognized as old [hits] vs. old images incorrectly judged as new [misses]) on spectral power, at any frequencies, at any channels, at any time.*

We followed the same analysis approach described in section 3.3, this time on *old-hit/old-miss* conditions and including all frequencies, timepoints, and channels.

3.6 RQ4a

The *EEGManyPipelines* team asked us to test the following:

> *RQ4a. There are effects of subsequent memory (i.e., a difference between images that will be successfully remembered vs. forgotten on a subsequent repetition) on EEG voltage at any channels, at any time.*

We followed the same analysis approach described in section 3.4, comparing *remembered/forgotten* conditions and including all timepoints and channels.

3.7 RQ4b

The *EEGManyPipelines* team asked us to test the following:

> *RQ4b. There are effects of subsequent memory (i.e., a difference between images that will be successfully remembered vs. forgotten on a subsequent repetition) on spectral power, at any frequencies, at any channels, at any time.*

We followed the same analysis approach described in section 3.5, this time on *remembered/forgotten* conditions and including all frequencies, timepoints, and channels.

Table 1

RQ1. *Means, 95% credible intervals (CI), and $\hat{R}$ statistic of the posterior distributions of intercept and β, separately for the models with informative, weakly informative, and uninformative priors (see section 3.1).*

Prior	Parameter	Mean	CI_low	CI_high	Rhat
informative	intercept	4.62	3.17	6.02	1.01
informative	beta	-0.32	-0.56	-0.09	1.00
weakly_informative	intercept	4.68	3.22	6.28	1.01
weakly_informative	beta	-0.33	-0.57	-0.09	1.00
uninformative	intercept	4.73	3.28	6.13	1.00
uninformative	beta	-0.33	-0.58	-0.10	1.00

Note. The slope refers to manmade minus natural conditions.

4 Results

4.1 RQ1

4.1.1 Descriptives and diagnostics. Table 1 shows descriptives (means and 95% CI) of the posterior distributions of intercept and β. Model diagnostics (see $\hat{R}$ in Table 1 and Figure 7) confirmed that all models successfully converged.

As anticipated in section 3.1, posterior distributions did not remarkably differ as a function of prior. Therefore, all subsequent results refer to the model with informative priors.

4.1.2 Hypothesis testing. The results of the **HDI + ROPE** procedure showed that 95% of the posterior distribution of the N1 amplitude difference between *manmade* and *natural* scenes was outside of a region of practical equivalence up until $\pm 0.08\ \mu V$. In

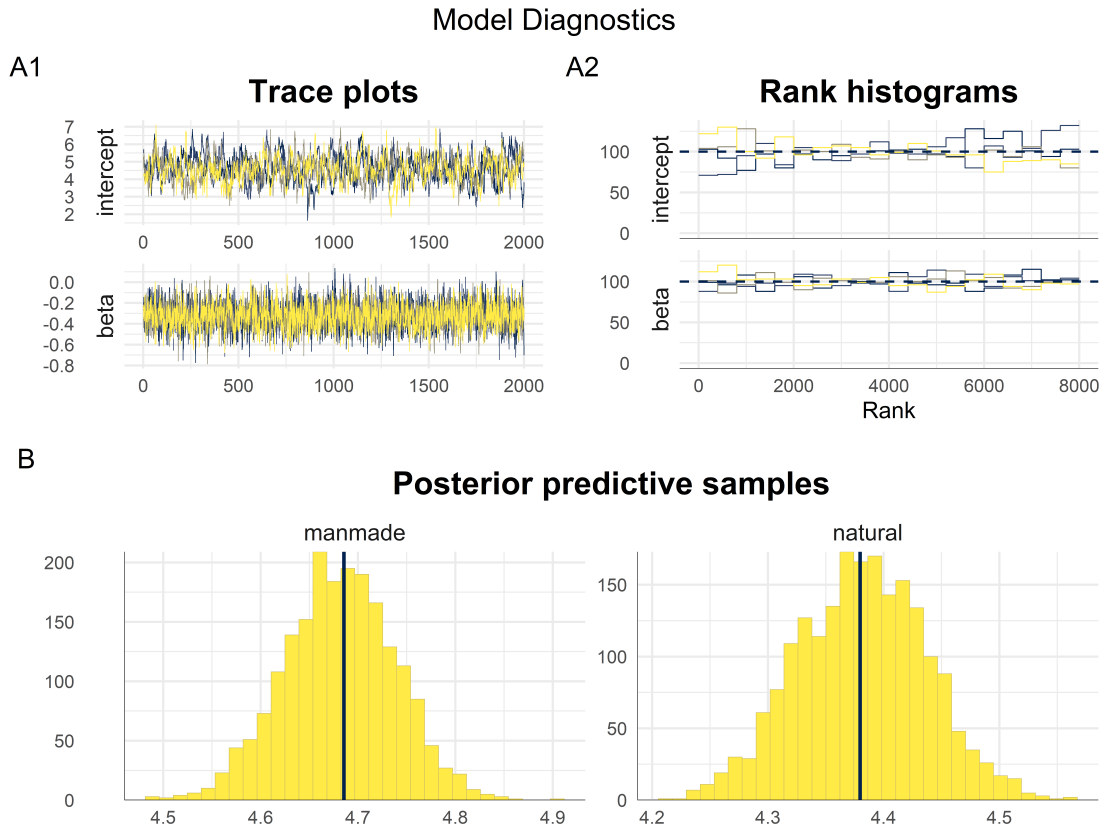

Figure 7. **RQ1**. Trace plots (panel *A1*), rank plots (panel *A2*), and graphical posterior predictive checks (panel *B*) for the model with informative priors.

other words, ***manmade* scenes elicited an N1 whose amplitude was at most 0.08** μV larger than the N1 elicited by *natural* scenes (see Figure 8).

4.2 RQ2a

4.2.1 Descriptives and diagnostics. Table 2 shows descriptives (means and 95% CI) of the posterior distributions of intercept and β. Model diagnostics (see $\hat{R}$ in Table 2 and Figure 9) confirmed that all models successfully converged.

As anticipated in section 3.1, posterior distributions did not remarkably differ as a function of prior. Therefore, all subsequent results refer to the model with informative priors.

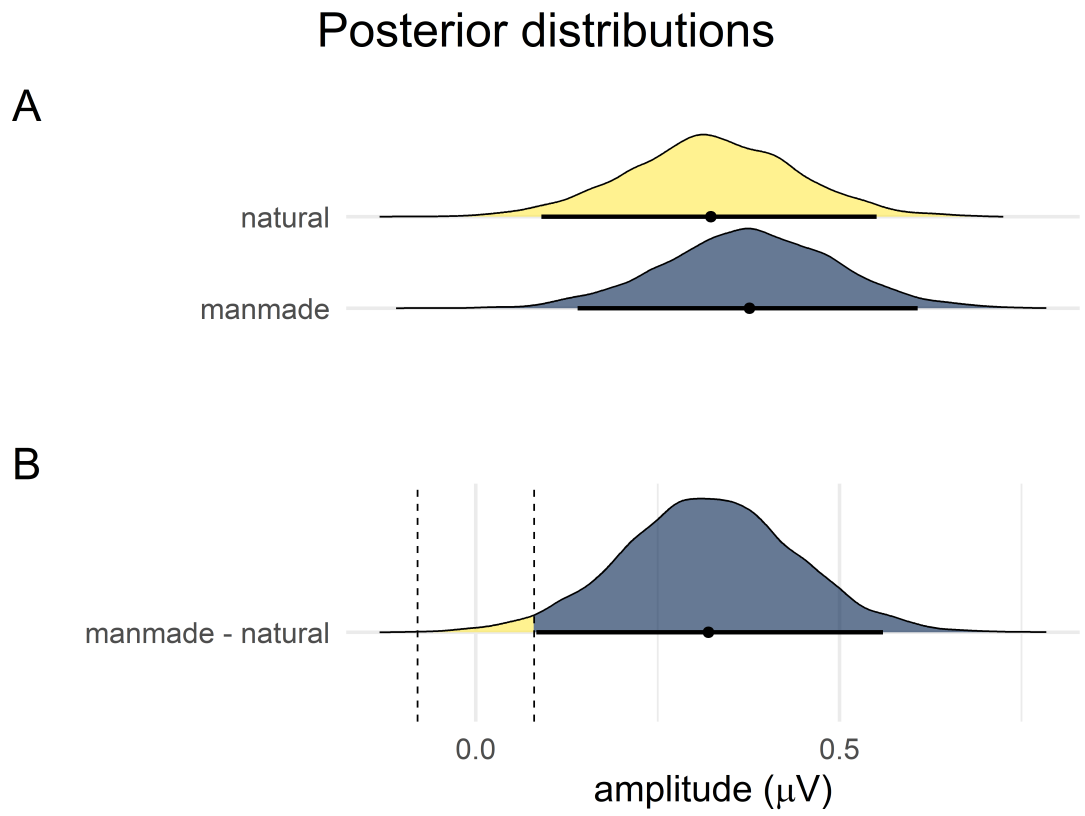

Figure 8. Panel *A* shows the posterior distributions of *manmade* and *natural* conditions. Panel *B* shows the posterior distribution of their difference (*manmade* minus *natural*), as well as the largest ROPE ($\pm 0.08\ \mu V$) for which the null hypothesis was rejected by using the *HDI + ROPE* procedure described in section 3.1.

4.2.2 Hypothesis testing. The results of the **HDI + ROPE** procedure showed that 95% of the posterior distribution of the amplitude difference between *old* and *new* scenes was outside of a region of practical equivalence up until $\pm 0.29\ \mu V$. In other words, ***new* scenes elicited electrophysiological brain activity whose amplitude was at most 0.29** μV more negative than the activity elicited by *old* scenes (see Figure 10).

Table 2

***RQ2a**. Means, 95% credible intervals (CI), and $\hat{R}$ statistic of the posterior distributions of intercept and β, separately for the models with informative, weakly informative, and uninformative priors (see section 3.1).*

Prior	Parameter	Mean	CI_low	CI_high	Rhat
informative	intercept	-6.63	-7.87	-5.35	1.01
informative	beta	0.44	0.29	0.59	1.00
weakly_informative	intercept	-6.51	-7.89	-5.27	1.02
weakly_informative	beta	0.44	0.29	0.59	1.00
uninformative	intercept	-6.46	-7.82	-5.19	1.01
uninformative	beta	0.44	0.29	0.58	1.00

Note. The slope refers to manmade minus natural conditions.

4.3 RQ2b, RQ2c, RQ3a, RQ3b, RQ4a, RQ4b

No statistically significant clusters were identified for any of these research questions, ***at any time point, electrode, or frequency***.

5 Software

EEG preprocessing was carried out using `MNE-Python` *v*0.24.1 (Gramfort, 2013) in *Python* *v*3.9.7 (Van Rossum & Drake, 2009) and *Spyder IDE* *v*5.1.5 (Raybaut, 2009). Analysis, visualization, and report generation were carried out in *R* *v*4.1.3 (R Core Team, 2022) and *RStudio IDE* *v*2022.02.1+461 (RStudio Team, 2020). We used the following *R* packages (and their dependencies):

- **data wrangling and analysis**: `here` *v*1.0.1 (Müller, 2020), `Rmisc` *v*1.5 (Hope,

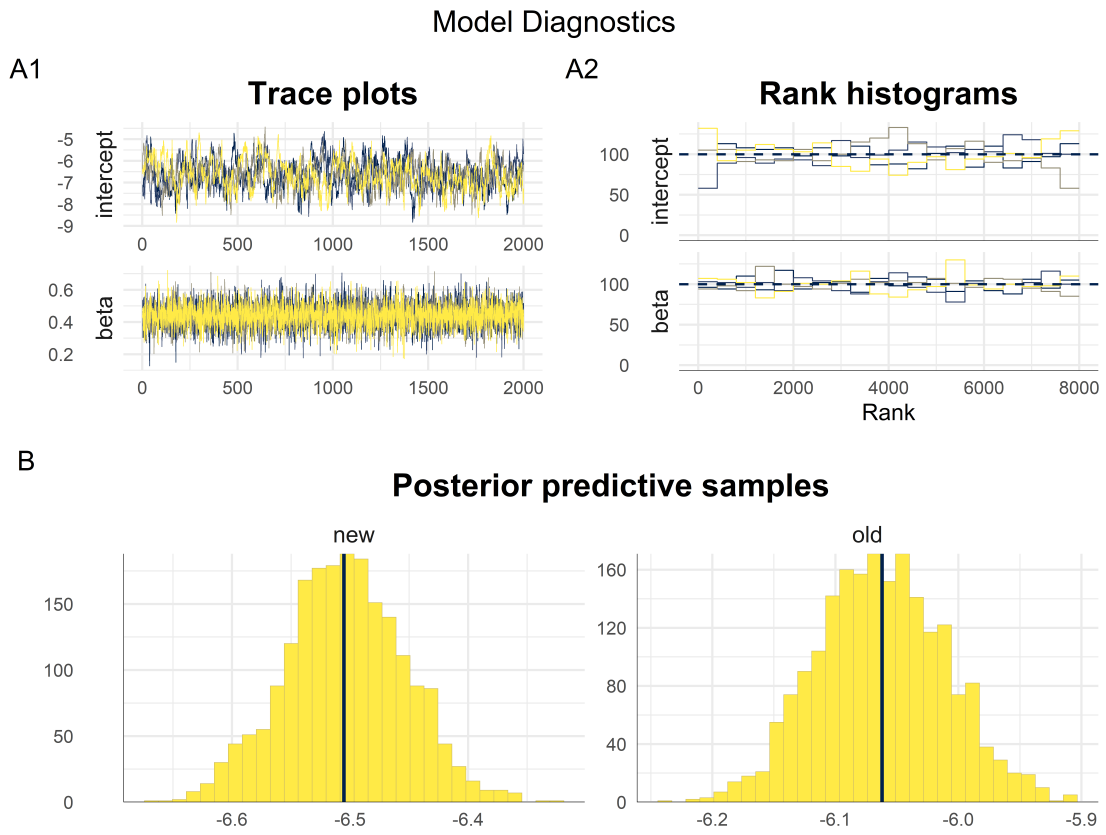

Figure 9. **RQ2a**. Trace plots (panel *A1*), rank plots (panel *A2*), and graphical posterior predictive checks (panel *B*) for the model with informative priors.

2022), `tidyverse` *v*1.3.1 (Wickham et al., 2019) – in particular `tibble` *v*3.1.6 (Müller & Wickham, 2022), `tidyr` *v*1.2.0 (Wickham & Girlich, 2022), `readr` *v*2.1.2 (Wickham, Hester, & Bryan, 2022), `dplyr` *v*1.0.9 (Wickham, François, Henry, & Müller, 2022) –, `brms` *v*2.17.0 (Bürkner, 2018), `eegUtils` *v*0.7.0 (Craddock, 2022), `emmeans` *v*1.7.3 (Lenth, 2022), `bayestestR` *v*0.11.5.1 (Makowski, Ben-Shachar, & Lüdecke, 2019)

- **visualization**: `ggplot2` *v*3.3.6 (Wickham, 2016), `eegUtils` *v*0.7.0 (Craddock, 2022), `bayesplot` *v*1.9.0 (Gabry & Mahr, 2022), `viridis` *v*0.6.2 (Garnier et al., 2021-04-11, 2021-04), `tidybayes` *v*3.0.2 (Kay, 2022), `patchwork` *v*1.1.1 (Pedersen, 2020)

- **report generation**: `knitr` *v*1.39 (Xie, 2022), `rmarkdown` *v*2.14 (Allaire et al., 2022),

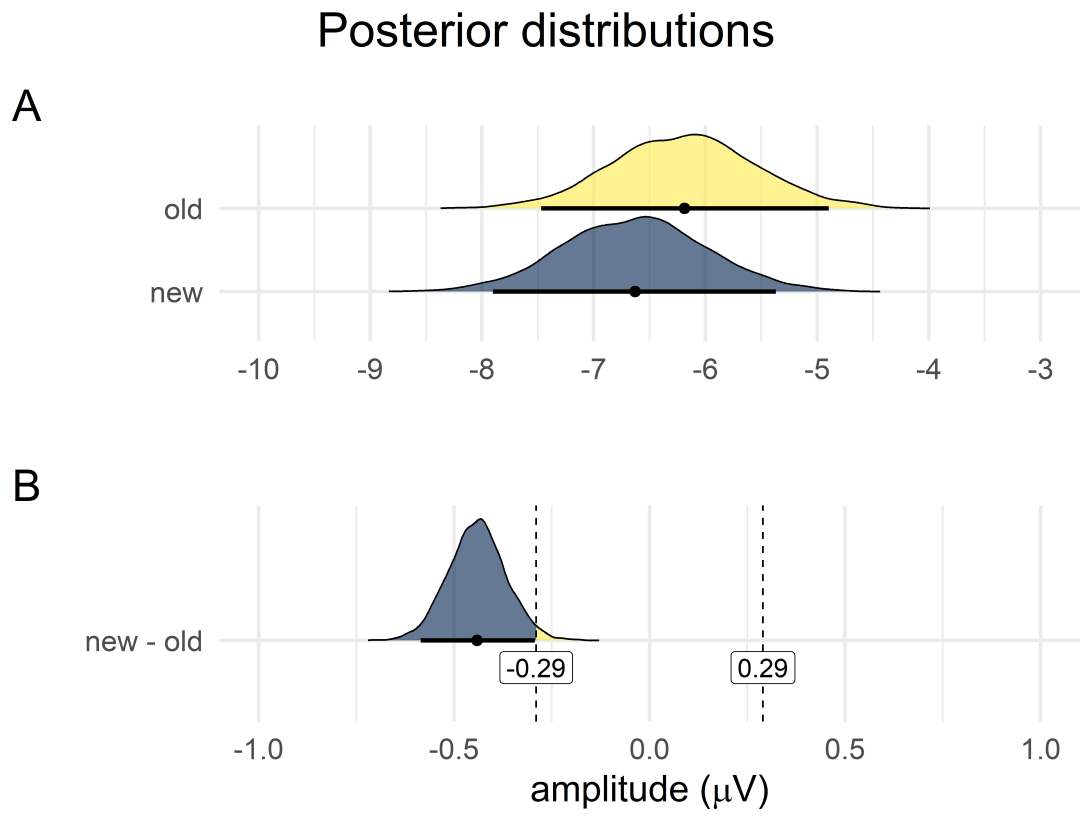

Figure 10. **RQ2a**. Panel *A* shows the posterior distributions of *new* and *old* conditions. Panel *B* shows the posterior distribution of their difference (*new* minus *old*), as well as the largest ROPE ($\pm 0.29\ \mu V$) for which the null hypothesis was rejected by using the *HDI + ROPE* procedure described in section 3.1.

`papaja` *v*0.1.0.9999 (Aust & Barth, 2022)

6 ANA COULD YOU PLEASE ADD HERE INFORMATION ABOUT MAKE AND THE R PACKAGES YOU USED FOR REPRODUCIBILITY (E.G., RETICULATE)? NO NEED TO ADD REFERENCES VIA RSTUDIO, BUT IF YOU WRITE DOWN DOIs YOU'LL MAKE MY LIFE EASIER :)

7 References

Ablin, P., Cardoso, J.-F., & Gramfort, A. (2018). Faster independent component analysis by preconditioning with hessian approximations. *IEEE Transactions on Signal Processing*, *66*(15), 4040–4049. https://doi.org/10.1109/tsp.2018.2844203

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Iannone, R. (2022). *Rmarkdown: Dynamic documents for R.*

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., Langen, J. van, & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Aust, F., & Barth, M. (2022). *Papaja: Prepare reproducible APA journal articles with R Markdown.*

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/rj-2018-017

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities. *American Journal of EEG Technology*, *25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163

Craddock, M. (2022). *eegUtils: Utilities for electroencephalographic (EEG) analysis.*

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*(2), 240–261. https://doi.org/10.1037/met0000065

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for

psychological research. *Psychological Review*, *70*(3), 193–242. https://doi.org/10.1037/h0044139

Fischler, M. A., & Bolles, R. C. (1987). *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. Elsevier. https://doi.org/10.1016/b978-0-08-051581-6.50070-2

Gabry, J., & Mahr, T. (2022). *Bayesplot: Plotting for bayesian models*.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Garnier, S., Ross, N., BoB Rudis, Filipovic-Pierucci, A., Galili, T., Timelyportfolio, ... JJ Chen. (2021-04-11, 2021-04). *Sjmgarnier/viridis: Viridis 0.6.0 (pre-CRAN release)*. Zenodo. https://doi.org/10.5281/ZENODO.4679424

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Homan, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*, *15*(1), 1593–1623.

Hope, R. M. (2022). *Rmisc: Ryan miscellaneous*.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Kay, M. (2022). *Tidybayes: Tidy data and geoms for bayesian models*. https://doi.org/10.5281/zenodo.1308151

Kruschke, J. K. (2018). Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

https://doi.org/10.1177/2515245918771304

Lenth, R. V. (2022). *Emmeans: Estimated marginal means, aka least-squares means.*

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). *bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. 4*, 1541. https://doi.org/10.21105/joss.01541

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Müller, K. (2020). *Here: A simpler way to find your files.*

Müller, K., & Wickham, H. (2022). *Tibble: Simple data frames.*

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242. https://doi.org/10.1044/2018_jslhr-s-18-0006

Natarajan, R., & Kass, R. E. (2000). Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, *95*(449), 227–237. https://doi.org/10.1080/01621459.2000.10473916

Pedersen, T. L. (2020). *Patchwork: The composer of plots.*

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

R Core Team. (2022). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria.

Raybaut, P. (2009). Spyder-documentation. *Available Online at: Pythonhosted. Org*.

RStudio Team. (2020). *RStudio: Integrated development environment for R* [Manual]. Boston, MA.

Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). *Welcome to the tidyverse*. *4*, 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*.

Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*.

Wickham, H., Hester, J., & Bryan, J. (2022). *Readr: Read rectangular text data*.

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Xie, Y. (2022). *Knitr: A general-purpose package for dynamic report generation in R*.