EEGmanypipelines

Julian Q. Kosciessa*[1][2]

[1]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin and London; [2]Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.

* Email: kosciessa@mpib-berlin.mpg.de; https://orcid.org/0000-0002-4553-2794

1. Introduction

The objective of the data analysis is to test the following hypotheses:

1. 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.
2. 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 ...
   a.) on EEG voltage at fronto-central channels.
   b.) on theta power at fronto-central channels.
   c.) on alpha power at posterior channels.
3. 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]) ...
   a.) on EEG voltage at any channels, at any time.
   b.) on spectral power, at any frequencies, at any channels, at any time.
4. There are effects of subsequent memory (i.e., a difference between images that will be successfully remembered vs. forgotten on a subsequent repetition) ...
   a.) on EEG voltage at any channels, at any time.
   b.) on spectral power, at any frequencies, at any channels, at any time.

2. Results

1.1 Effect of scene category on N1 amplitude

The first analysis assessed potential differences in visual N1 amplitude as a function of scene category, i.e., whether the presented image reflected a natural or manmade scene. To focus on evoked potentials, we subtracted single-trial baselines defined as the average voltage during the 200 ms preceding stimulus presentation. Then, the channel with the minimum voltage between 40 and 120 ms following stimulus presentation was identified. This analysis identified channel Oz, which we used in the subsequent analyses. Notably, inspection of individual ERPs indicated striking, and highly systematic N1 latency differences between two groups of participants: the initial 17 participants that were recorded in an electrically shielded booth showed a comparatively earlier visual N1 peak than participants that were recorded in a room without shielding. These systematic differences are unlikely to have arisen during preprocessing, as the preprocessing setup was identical across all subjects. Given that these two subgroups are differentiated by their systematic differences in shielding during recording, differences may be related to covarying factors such as the exact monitor and trigger setup. The sparse information provided about the setup does not allow clear conclusions at this point, but potential confounding factors should be reconsidered. To account for this observed variability in N1 latency, N1 peak amplitude was measured at the first local minimum between 40 and 120 ms of the across-condition average Oz trace. A subsequent paired-sample t-test did not indicate statistically significant differences in N1 peak amplitudes between conditions (t(32)= 1.097964244609333, p = 0.280413264498449) at a canonical 5% significance threshold. This analysis does therefore not provide evidence that scene category modulates visual N1 peak amplitude.

Hypothesis 1 was **not** confirmed.
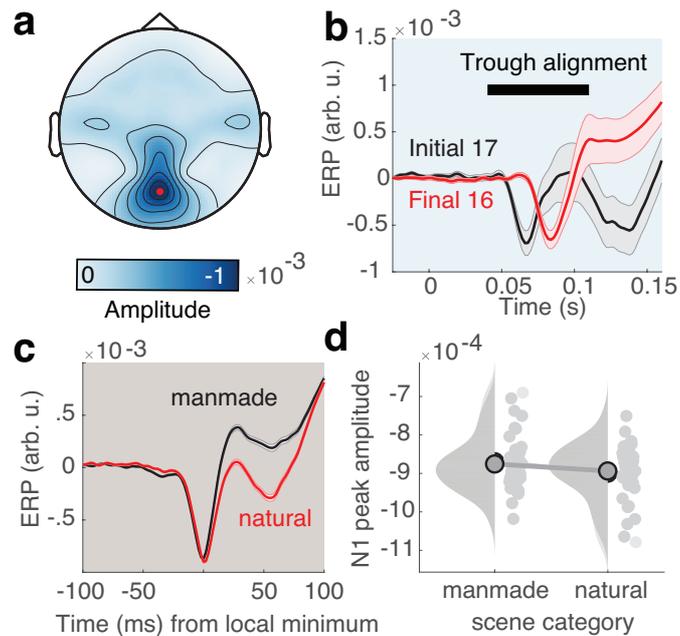
Figure 1. N1 peak amplitude is not modulated by scene category. (a) The channel showing minimum voltage amplitude between 40 to 120 ms following stimulus onset was used to extract the N1. The red dot indicates the analyzed sensor Oz. (b) Data isnpection revealed systematic differences in N1 latency between two participant subgroups. Traces are mean +/- S.E.M. (c) Event-related potential at Oz after alignment to individual N1 peak latency. Traces are mean +/- within-subject S.E.M. (d) No differences in N1 peak amplitude as a function of scene category were observed. Data are within-subject centered to emphasize potential between-condition differences.

1.2 Effect of image novelty

a. on fronto-central EEG voltage

Next, we tested whether image novelty affected EEG voltage, i.e., whether voltage differed depending on whether the presented image was a repeat ("old") or presented for the first time ("new"). We performed a task PLS to test for potential differences between conditions. Task PLS assesses optimal statistical relations between time-frequency matrices and experimental conditions (here two: old and new stimuli). We included only time points during the specified 300 to 500 ms post-stimulus time window, and set data outside the hypothesized locations to zero (see also methods). This task PLS indicated one significant latent variable (permuted p ~ 0). Prominent loadings were observed at a mediofrontal channel as well as a central cluster of channels (Figure 2a). At these locations, voltage was increased during old relative to new images during the 300 to 500 ms post-stimulus time window (Figure 2b-d).

Hypothesis 2a was confirmed.
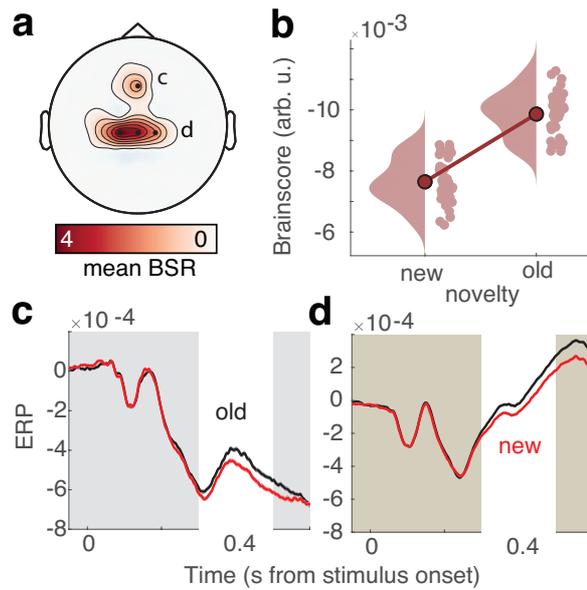
Figure 2. Repeat image presentations increase voltage over mediofrontal and central channels. (a) Max. bootstrap ratio between 300 and 500 ms following stimulus onset. (b) Multivariate brainscores indicated a stronger expression of this multivariate pattern for old compared to new images. (c, d) Traces extracted from frontal (c) and central (d) channels highlighted in panel a. Traces indicate means +/- within-subject SEMs.

b. on frontal theta power

Next, we performed a similar task PLS for theta power. This task PLS indicated significant differences of frontal theta power as a function of image novelty (*permuted* p ~ 0). Specifically, mediofrontal theta power increased when the presented image was a repeat compared to when it was presented for the first time, starting during the specified 300 to 500 ms post-stimulus time window.

Hypothesis 2b was confirmed.
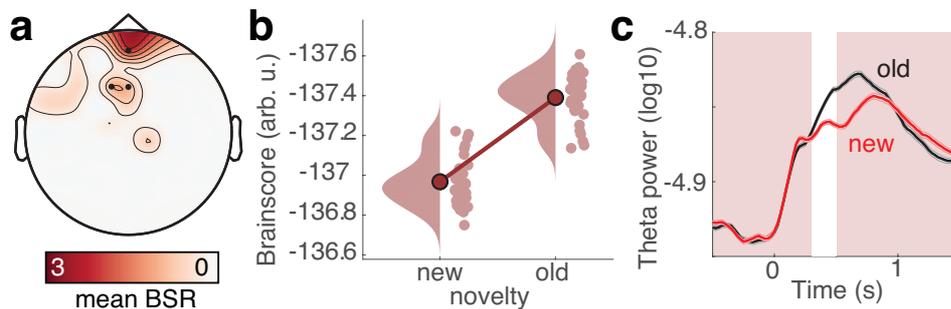
Figure 3. Theta power at frontal channels as a function of image novelty. (a) Topography of bootstrap ratios from the task PLS. (b) Brainscores indicated stronger expression of the multivariate pattern for old as compared to new images. (c) Theta power traces extracted from the most robust channels highlighted in panel a. Traces indicate means +/- within-subject SEMs.

c. on posterior alpha power

Next, we performed a task PLS that included data from posterior and occipital channels during the 300 to 500 ms post-stimulus time window. This task PLS did not indicate significant differences of posterior alpha power as a function of image novelty (first latent variable: *permuted* p = 0.203796203796204). Figure 4 shows alpha traces as a function of old/new image presentation.

Hypothesis 2c was **not** confirmed.
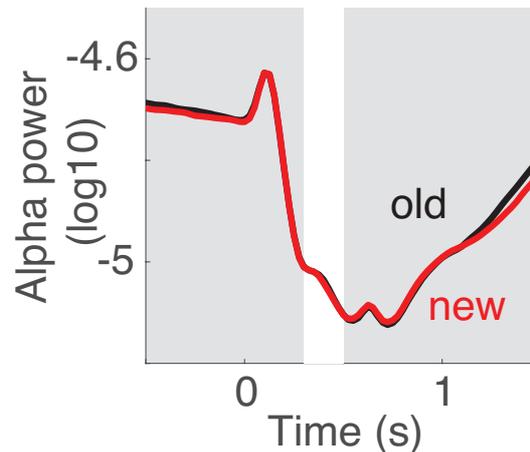
Figure 4. Alpha power at posterior-occipital channels does not change as a function of image novelty. Traces are averaged across all channels with labels starting with 'P', or 'O'. The white shading indicates the targeted time period. Traces indicate means +/- within-subject SEMs.

1.3 Effect of successful recognition

a. on EEG voltage

An exploratory task PLS of voltage changes as a function of successful recognition (i.e., a difference between old images correctly recognized as old [hits] vs. old images incorrectly judged as new [misses]) indicated a significant latent variable (*permuted* p ~ 0). Prominent differences were observed at multiple channels and time periods (Figure 5). During an early time window from around 300 to 500 ms following stimulus presentation, mediofrontal voltage was elevated for subsequent hits compared to misses, while right anterior voltage was decreased (Figure 5c, top). During a later time window from around 600 to 1000 ms, centroparietal voltage was increased for hits compared to misses, while right anterior and posterior voltage was decreased (Figure 5c, middle). Finally, during a late time window from around 1.2 to 1.2 s, bilateral centro-posterior voltage was decreased for subsequent hits, while lateral and posterior voltage showed relative increases (Figure 5c, bottom). Taken together, this exploratory analysis indicates that EEG voltage shows diverse differences depending on whether a presented 'old' stimulus was correctly recognized as old or not.
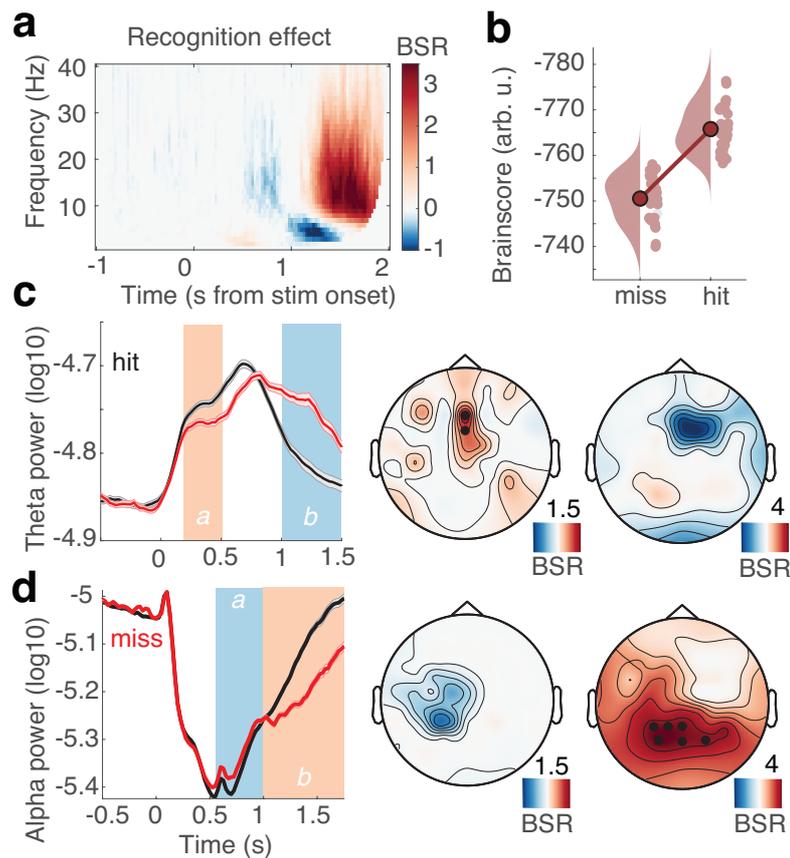
Hypothesis 3a was confirmed.
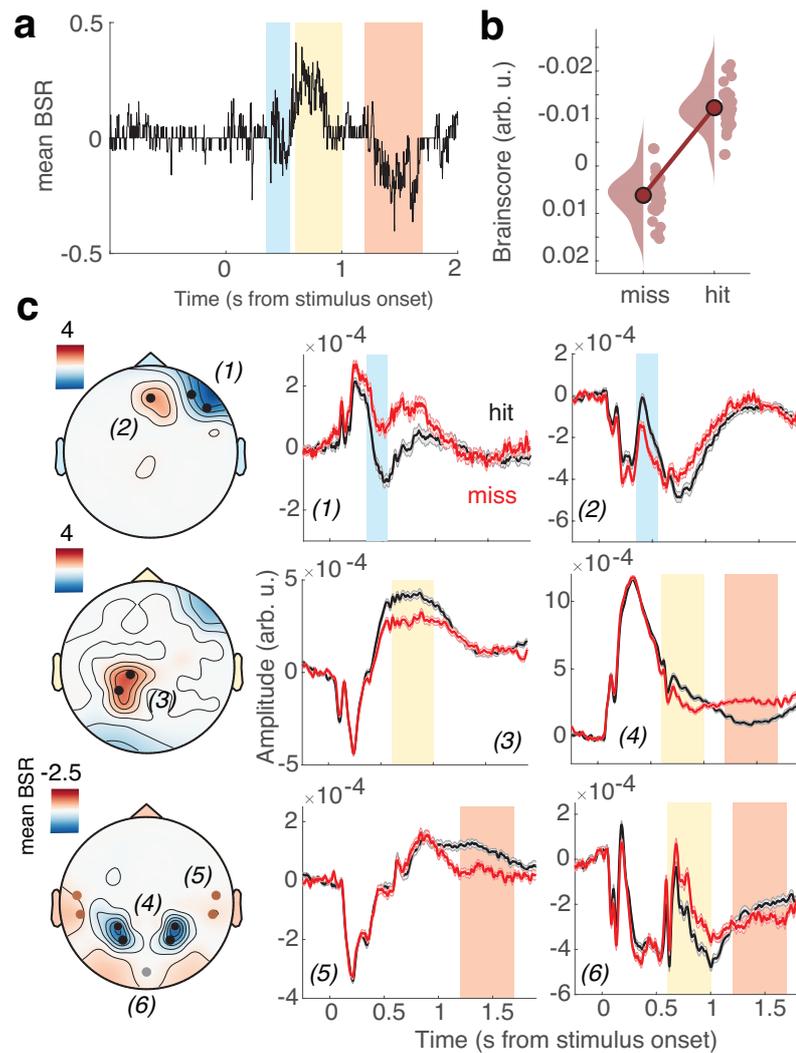
Figure 5. Effect of successful recognition of repeated images on EEG voltage. (a) Average bootstrap ratios (BSRs) indicated robust differences at multiple timepoints. (b) Expression of this pattern was more pronounced in hit as compared to miss trials. (c) Topographies showing mean BSRs across the time windows shown via colored shading in a. Traces have been extracted from the channels that are indicated and numbered in the topographies. Identical colored shading is provided on top of traces for quick reference. Traces indicate means +/- within-subject SEMs.

b. **on spectral power**

An exploratory task PLS of spectral power changes as a function of successful recognition indicated a significant latent variable (*permuted* p ~ 0). This latent variable reflected spectral power differences between images that were correctly identified as having previously been presented (hits) or incorrectly identified as novel (misses). These differences were observed at multiple frequencies and time windows (Figure 6). Relative to misses, hits showed initially larger mediofrontal theta power, followed by decreased theta power later during the trial (Figure 6c). The opposite pattern was observed for posterior alpha power, with initially decreased alpha power for hits relative to misses during an earlier time window around 500 to 1000 ms following stimulus onset being followed by decreased theta power later during the trial (Figure 6d).

Hypothesis 3b was confirmed.

Figure 6. Effects of successful recognition of repeated images on spectral power. (a) Average bootstrap ratios (BSRs) indicated that multiple timepoints and frequencies exhibited robust differences. (b) Expression of this pattern was more pronounced in hit as compared to miss trials. (c, d) Traces and topographies of theta and alpha power. Traces indicate means +/- within-subject SEMs and were extracted from the channels indicated on the right. Colored shading indicates the time periods across which bootstrap ratios were averaged for the topographies.

1.4 Effect of subsequent memory

a. on EEG voltage

An exploratory task PLS of EEG voltage differences as a function of subsequent memory i.e., a difference between images that will be successfully remembered vs. forgotten on a subsequent repetition) indicated one significant latent variable (*permuted* p = 0.035964035964036). However, sub-001 represented a notable outlier in this analysis. No significant latent variable (indicating condition differences) was observed after removing this outlier prior to analysis (1st LV: *permuted* p: 0.0719280719280719). The analysis therefore does not indicate robust differences in EEG voltage as a function of subsequent memory

Hypothesis 4a was **not** confirmed.

b. on spectral power

An exploratory task PLS of spectral power changes as a function of subsequent memory for the presented images indicated a significant latent variable (*permuted* p = 0.002997002997003). This variable primarily reflected differences in frontal theta and posterior alpha power (Figure 7).

Specifically, frontal theta power was larger for subsequently remembered than forgotten stimuli during approx. the first second following stimulus presentation. Moreover, posterior alpha power was more strongly reduced for subsequently remembered than forgotten stimuli from around 800 to 1000 ms following stimulus onset.

Hypothesis 4a was confirmed.
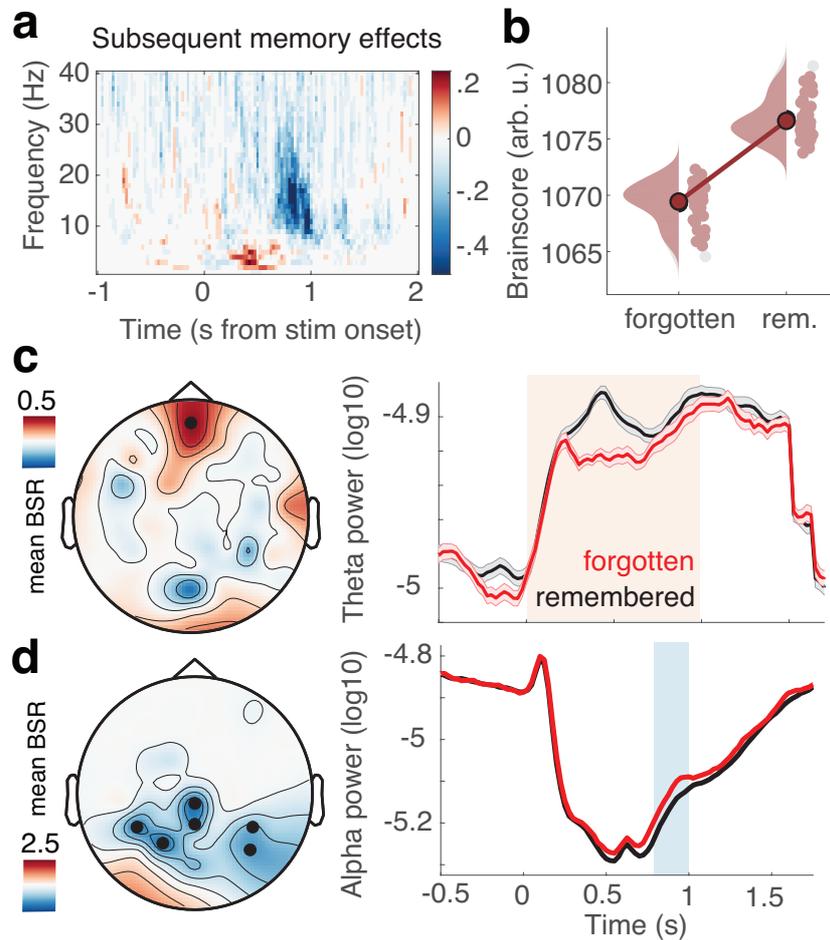
Figure 7. Effects of subsequent memory on spectral power. (a) Average bootstrap ratios (BSRs) dominantly indicated earlier differences in theta frequencies, and later differences in alpha power. (b) Expression of this pattern was more pronounced when stimuli were remembered rather than forgotten. (c, d) Topographies of average bootstrap ratios in the theta (c) and alpha (d) frequency range, alongside traces of spectral power. Traces indicate means +/- within-subject SEMs and were extracted from the channels indicated on the right. Colored shading indicates the time periods across which bootstrap ratios were averaged for the topographies.

3. Methods

Preprocessing. Preprocessing and analysis of EEG data were conducted with the FieldTrip toolbox (revision 021ac9105) (Oostenveld, Fries, Maris, & Schoffelen, 2011) and using custom-written MATLAB (The MathWorks Inc., Natick, MA, USA) code. EEG data were filtered using a 4th order Butterworth filter with a pass-band of 1 to 100 Hz. Subsequently, data were downsampled to 250 Hz and all channels were re-referenced to mathematically averaged mastoids. Blink, movement and heart-beat artifacts were identified using Independent Component Analysis (ICA) (Bell & Sejnowski, 1995) and removed from the signal. Prior to ICA detection, large artifactual time periods were manually labelled and excluded from the ICA to facilitate the detection of more specific components. Artifact-contaminated channels (determined across epochs) were automatically detected using (a) an adapted version of the FASTER algorithm (Nolan, Whelan, & Reilly, 2010), and by (b) detecting outliers exceeding three standard deviations of the kurtosis of the distribution of power values in each epoch within low (0.2-2 Hz) or high (30-100 Hz) frequency bands, respectively. Rejected channels were interpolated using spherical splines (Perrin, Pernier, Bertrand, & Echallier, 1989). Subsequently, noisy epochs were likewise excluded based on a custom implementation of FASTER and on recursive outlier detection. Finally, we epoched the cleaned data from -1s to 2s surrounding the onset of visual stimuli. Given that the latter half of the sample was recorded without electrical shielding, line noise at 50, 100 and 150 Hz was removed using Discrete Fourier Transform as implemented in FieldTrip. To enhance spatial specificity, scalp current density estimates were derived via 4th order spherical splines [82] using a standard 1005 channel layout (conductivity: 0.33 S/m; regularization: 1^-05; 14th degree polynomials).

ERP analyses. We baseline-adjusted signals by subtracting the average single-trial voltage during the 200 ms window preceding stimulus onset. To identify the channel with a maximal representation of the early N1 potential, the minimum voltage between 40 and 120 ms following stimulus presentation was identified. Given that inspection of individual data indicated systematic N1 latency differences between different subgroups (see Figure 1b), N1 peak amplitude was measured at the first local minimum between 40 and 120 ms of the across-condition average Oz trace. Due to the low-dimensional nature of the data, a two-sided paired t-test was performed to test for potential differences as a function of scene category.

Spectral power. Time frequency transformation was performed using superlets (Moca, Barzan, Nagy-Dabacan, & Muresan, 2021) for linearly spaced frequencies between 1 and 40 Hz (step size: 1 Hz). Superlets reflect a frequency-wise combination of Morlet wavelets of varying cycle widths, providing a frequency-adaptive temporal resolution of spectral power estimates. This analysis used a base wavelet of 3 cycles, with additive combination of cycle lengths defining a superlet, as implemented in FieldTrip.

Statistical analyses. All statistical tests used a canonical two-sided 5% criterion to assess significance of effects. Individual outliers were identified when condition differences exceeded three scaled median absolute standard deviations, and are reported in light grey in the RainCloud (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019) distribution plots. Where a p-value of approx. zero is reported, no more precise estimate could be obtained.

Multivariate partial least squares analyses. For data with a high-dimensional structure, we performed multivariate partial least squares (PLS) analyses (Krishnan, Williams, McIntosh, & Abdi, 2011; McIntosh, Bookstein, Haxby, & Grady, 1996), specifically 'task PLS'. Task PLS begins by calculating a between-subject covariance matrix (COV) between conditions and each neural value (e.g., time-space-frequency power), which is then decomposed using singular value decomposition (SVD). This yields a left singular vector of experimental condition weights (U), a right singular vector of brain weights (V), and a diagonal matrix of singular values (S). Task PLS produces orthogonal latent variables (LVs) that reflect optimal relations between experimental conditions and the neural data. For each LV (ordered strongest to weakest in S), a data pattern results which depicts the strongest available relation between brain data and experimental conditions. Significance of detected relations was assessed using 1000 permutation tests of the singular value corresponding to the LV. A subsequent bootstrapping procedure indicated the robustness of within-LV neural saliences across 1000 resamples of the data (Efron & Tibshirani, 1986). By dividing each brain weight (from V) by its bootstrapped standard error, we obtained "bootstrap ratios" (BSRs) as normalized robustness estimates. We generally thresholded BSRs at values of ±3.00 (~99.9% confidence interval). Note that for visualization, a threshold BSR of 3 is initially applied to the input matrix, setting all subthreshold

data to zero. Due to this convention, average BSRs are sensitive to the relative extent of super-threshold data across the averaged dimension. We also obtained a summary measure of each participant's robust expression of a particular LV's pattern (a within-person "brain score") by multiplying the vector of brain weights (V) from each LV by each participant's vector of neural values (P), producing a single within-subject value: Brain score $= VP'$.

Testing specific spatio-temporal hypotheses within PLS models. When specific time periods were tested, only data falling within the respective time window were included in the PLS. Similarly, where hypotheses specified sensor locations, we constrained the input data to those sensors. For analyses targeting fronto-central voltage and theta power, any channel whose labels did not start with 'A', 'F', or 'C' were set to zero. For analyses targeting posterior alpha power, any channel whose labels did not start with 'P', or 'O' were set to zero. For hypotheses that were unspecific to the approximate timing of potential changes, we included the epochs ranging from -1s to 2 s around stimulus onset. It is possible that this time window also includes responses and potential condition differences therein. The provided data did not contain single-trial RTs to allow for a closer inspection of response-aligned signals however.

Within-subject centering. To visually emphasize condition differences, we use within-subject centering across repeated measures conditions by subtracting individual cross-condition means and adding global group means. For these visualizations, only the mean of the dependent values directly reflects the original units of measurement, as individual data points by construction do not reflect between-subject variation averaged across conditions. This procedure equals the creation of within-subject standard errors (Loftus & Masson, 1994). Within-subject centering is exclusively used for display and is noted in legends.

Data and script availability. All used code and data are available as DataLad (Halchenko et al., 2021) datasets, separated into the raw BIDS data (https://gin.g-node.org/juliankosciessa/eegmp_data), preprocessing code and data (https://gin.g-node.org/juliankosciessa/eegmp_preproc; doi: 10.12751/g-node.6nhr1y), and analysis code, data and figures (https://gin.g-node.org/juliankosciessa/eegmp_analysis; doi: 10.12751/g-node.98d6ks). Select code and data have been submitted in the format requested by the EEGmanypipelines committee.

4. References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res, 4*, 63. doi:10.12688/wellcomeopenres.15191.1

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129-1159. doi:10.1162/neco.1995.7.6.1129

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science, 1*(1), 54-75. doi:10.1214/ss/1177013815

Halchenko, Y., Meyer, K., Poldrack, B., Solanky, D., Wagner, A., Gors, J., . . . Hanke, M. (2021). DataLad: distributed system for joint management of code, data, and their relationship. *The Journal of Open Source Software, 6*(63), 3262. doi:10.21105/joss.03262

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage, 56*(2), 455-475. doi:10.1016/j.neuroimage.2010.07.034

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychon Bull Rev, 1*(4), 476-490. doi:10.3758/BF03210951

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage, 3*(3 Pt 1), 143-157. doi:10.1006/nimg.1996.0016

Moca, V. V., Barzan, H., Nagy-Dabacan, A., & Muresan, R. C. (2021). Time-frequency super-resolution with superlets. *Nat Commun, 12*(1), 337. doi:10.1038/s41467-020-20539-9

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods, 192*(1), 152-162. doi:10.1016/j.jneumeth.2010.07.015

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869. doi:10.1155/2011/156869

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. doi:10.1016/0013-4694(89)90180-6