

# Temporal dynamics of implicit moral evaluation: from empathy for pain to mentalizing processes

Kamela Cenka, Chiara Spaccasassi, Stella Petkovic, Rachele Pezzetta, Giorgio Arcara, Alessio Avenanti

## Supplementary material

### 1. Supplementary Methods

The dataset used in this study is part of a TMS-EEG experiment where we stimulated three different brain regions with TMS: the right TPJ for subjects 1 to 22, the right S1 for subjects 23 to 44 and left TPJ for subjects 45 to 66. To prevent potential cumulative effects on neural activity caused by repeated single pulse TMS, an inter-pulse interval of approximately 10 seconds was maintained between trials. Additionally, we segmented the EEG epochs and applied specific pre-processing steps to exclude any TMS-related artefacts (see in the main text, section 2.5: EEG preprocessing). However, to further ensure the validity of our ERP findings, we conducted additional analyses using Group (rTPJ, S1, lTPJ) as a between-subject factor to confirm the absence of ERP differences during the pre-TMS stimulation epochs.

#### 1.1. Analyses

##### 1.1.1. Cluster based approach

In line with our main ERPs analyses (see section 2.7 in main text: EEG analysis), we employed a whole-head cluster-based permutation test using independent samples t-statistics, analyzing data from 0 to 500 ms post-stimulus. Clusters were defined as groups of two or more neighboring electrodes where the t-statistic at a specific time point exceeded a p-value threshold of  $< 0.05$  (two-tailed independent samples t-test). We performed cluster-based permutation tests for each comparison based on 1000 random permutations, maintaining an alpha level below 0.05. This analysis was conducted across the three groups (rTPJ, lTPJ and S1) for each of the three conditions (IHS, AHS and NAS).

##### 1.1.2. Factorial ANOVA approach

To further explore potential between-groups differences within the six significant clusters identified in our main analyses (see section 3.1 in the main text: Cluster-Based Permutation Tests on ERPs), we extracted data corresponding to each significant cluster observed in the comparisons (IHS vs. NAS, AHS vs. NAS, and IHS vs. AHS). Specifically, for each comparison, we extracted data from the central and posterior clusters, along with their respective time windows. These data were then submitted to six mixed-factorial ANOVAs, using the pair of Conditions (e.g., IHS and NAS) as a within-subject variable and Group (rTPJ, S1, lTPJ,) as a between-subjects variable.

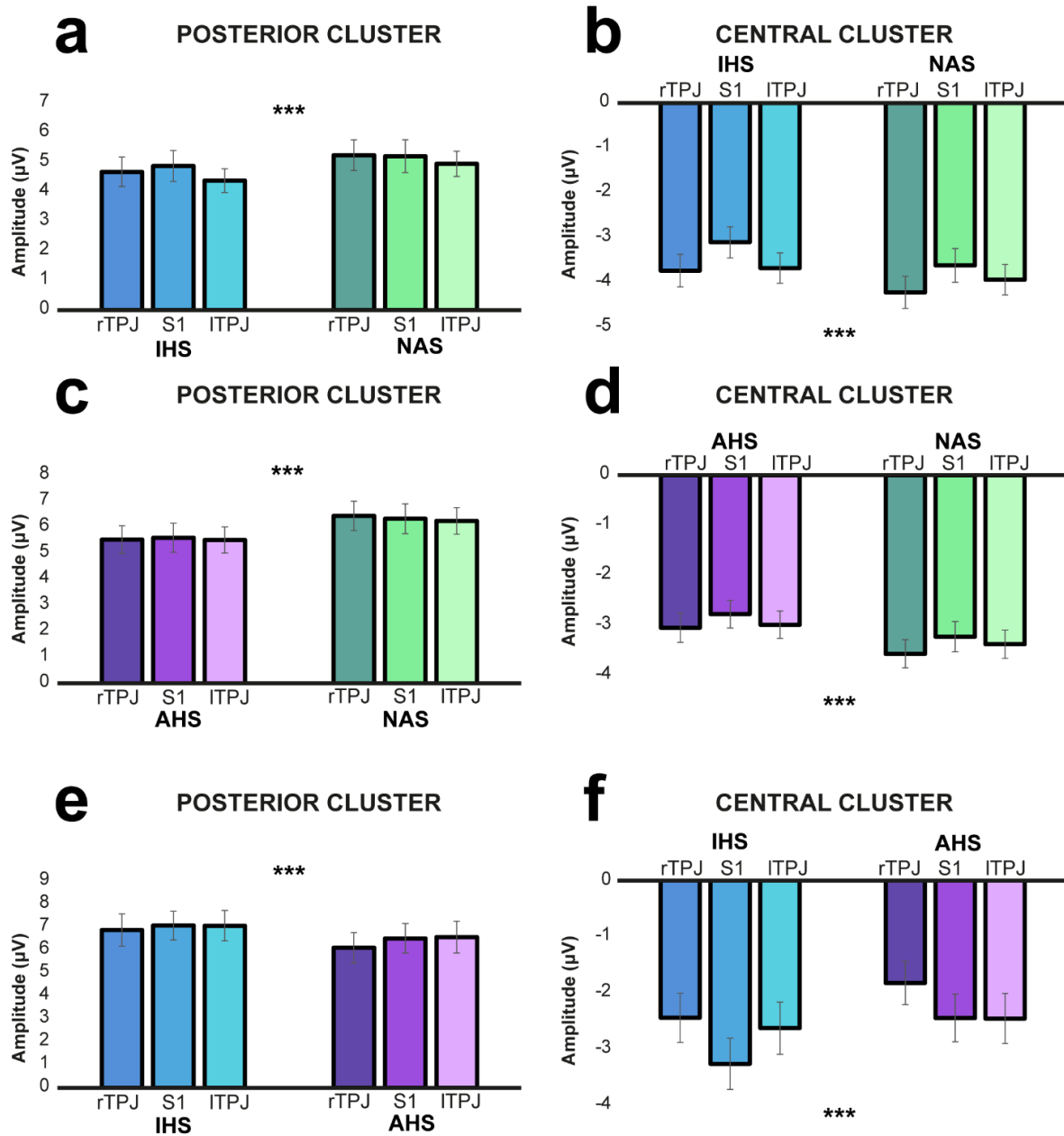
### 2. Supplementary Results

#### 2.1. Cluster based approach

The cluster-based analyses revealed no significant clusters among the groups (all  $p > 0.15$ ).

## 2.2. Factorial ANOVA approach

The mixed ANOVAs results revealed that while the main effect of Conditions was consistently significant (all  $F_{1,63} > 17.15$  and  $p < 0.001$ ), none of the interactions between Group and Conditions reached statistical significance, suggesting that the observed ERPs effects were consistent across the different groups (all  $F_{2,63} < 2.19$  and  $p > 0.12$ ).



**Figure S1. Between-Group ERP Analyses:** The bar graphs display the mean ERP amplitude for the posterior and central cluster across the three groups (rTPJ, S1, ITPJ) corresponding to IHS, AHS and NAS conditions. Panels (a) and (b) show the IHS vs. NAS comparison, where significant differences emerged in both the posterior (a) and central (b) clusters, with larger ERP amplitudes for NAS across all groups. Panels (c) and (d) display the AHS vs. NAS comparison, showing significant effects in both the posterior (c) and central (d) clusters, with larger ERP amplitudes for NAS. Panels (e) and (f) present the IHS vs. AHS comparison, highlighting significant differences in both the posterior (e) and central (f) clusters, where IHS elicited larger amplitudes compared to AHS. No significant Group x Condition interaction reached statistical significance, suggesting that the observed ERPs effects were consistent across the different groups (all  $F_{2,63} \leq 2.19$  and  $p \geq 0.12$ ). Error bars represent standard error of the mean, and significance levels are denoted by asterisks (\*\*\*)  $p < 0.001$ ).