

# SCIENTIFIC REPORTS



OPEN

## Machine learning provides novel neurophysiological features that predict performance to inhibit automated responses

Amirali Vahid<sup>1</sup>, Moritz Mückschel<sup>1</sup>, Andres Neuhaus<sup>2</sup>, Ann-Kathrin Stock<sup>1</sup>  & Christian Beste<sup>1</sup>

Neurophysiological features like event-related potentials (ERPs) have long been used to identify different cognitive sub-processes that may contribute to task performance. It has however remained unclear whether “classical” ERPs are truly the best reflection or even causal to observable variations in behavior. Here, we used a data-driven strategy to extract features from neurophysiological data of  $n = 240$  healthy young individuals who performed a Go/Nogo task and used machine learning methods in combination with source localization to identify the best predictors of inter-individual performance variations. Both Nogo-N2 and Nogo-P3 yielded predictions close to chance level, but a feature in between those two processes, associated with motor cortex activity (BA4), predicted group membership with up to ~68%. We further found two Nogo-associated features in the theta and alpha bands, that predicted behavioral performance with up to ~78%. Notably, the theta band feature contributed most to the prediction and occurred at the same time as the predictive ERP feature. Our approach provides a rigorous test for established neurophysiological correlates of response inhibition and suggests that other processes, which occur in between the Nogo-N2 and P3, might be of equal, if not even greater, importance.

Neurophysiological features such as event-related potentials (ERPs) correlate with behavioral performance in a wide range of different tasks and have therefore been used to identify different cognitive sub-processes that contribute to task performance. Yet still, it has remained largely unclear whether “classical” ERPs are truly the best reflection or even causal to observable variations in behavior. While there is a lot of literature that helps to interpret one’s findings<sup>1</sup>, most of the investigated differences / ERPs are defined by minima, maxima, or pre-defined time windows in the course of the neurophysiological times series. This excludes a wealth of data “in between” the typically quantified neurophysiological features. This bias is especially dramatic given that any given EEG signal is composed of different signals, which do not only vary in latency, but also stem from different neuronal generators within the human brain<sup>2-6</sup>. To address this issue, many recent studies have applied complex signal decomposition approaches<sup>2,7-9</sup>, but even those approaches usually limit themselves to minima and maxima of the different identified components and have therefore also remained correlational.

Fortunately, the recent rise in machine learning approaches and methods has equipped us with a new and powerful tool that can expediently and objectively identify differences in the entire EEG signal. It may furthermore help to identify neurophysiological features that allow to predict behavioral performance, instead of just correlating with it. In fact, it has recently been argued that machine learning approaches may foster theoretical achievements and improve our understanding of the brain dynamics underlying cognitive functions<sup>10</sup>. To achieve this objective, it is important to move beyond ERP components in EEG research<sup>10</sup>. While a few studies have used ERP data to predict behavioral performance with the help of support vector machines (SVMs)<sup>11-13</sup>, such approaches are still not widely used. Moreover, these approaches are often strongly influenced by assumptions about the functional relevance of specific ERP-components and therefore based on a biased selection of potential features<sup>11-13</sup>. This means that neurophysiological features currently used for behavioral predictions with SVM

<sup>1</sup>Cognitive Neurophysiology, Department of Child and Adolescent Psychiatry, Faculty of Medicine of the TU Dresden, Saxony, Germany. <sup>2</sup>Department of Psychiatry, Charite University Hospital Berlin, Berlin, Germany. Amirali Vahid, Moritz Mückschel, Ann-Kathrin Stock and Christian Beste contributed equally. Correspondence and requests for materials should be addressed to C.B. (email: [christian.beste@uniklinikum-dresden.de](mailto:christian.beste@uniklinikum-dresden.de))

approaches have often been pre-selected to represent certain neurophysiological characteristics, such as specific ERP peaks for which correlational differences have previously been observed. Even though these approaches can predict (i.e. classify) the difference between experimental conditions or correctly and incorrectly answered trials above chance level, classification performance is often not very high<sup>12,13</sup>. For a critical discussion of this issue, see<sup>14</sup>. In this context, please note that throughout the entire manuscript, the term prediction will be strictly used in the context provided by machine learning approaches. This is, how well can individuals with “good” or “bad” behavioral performance be distinguished with the help of a given neurophysiological feature?

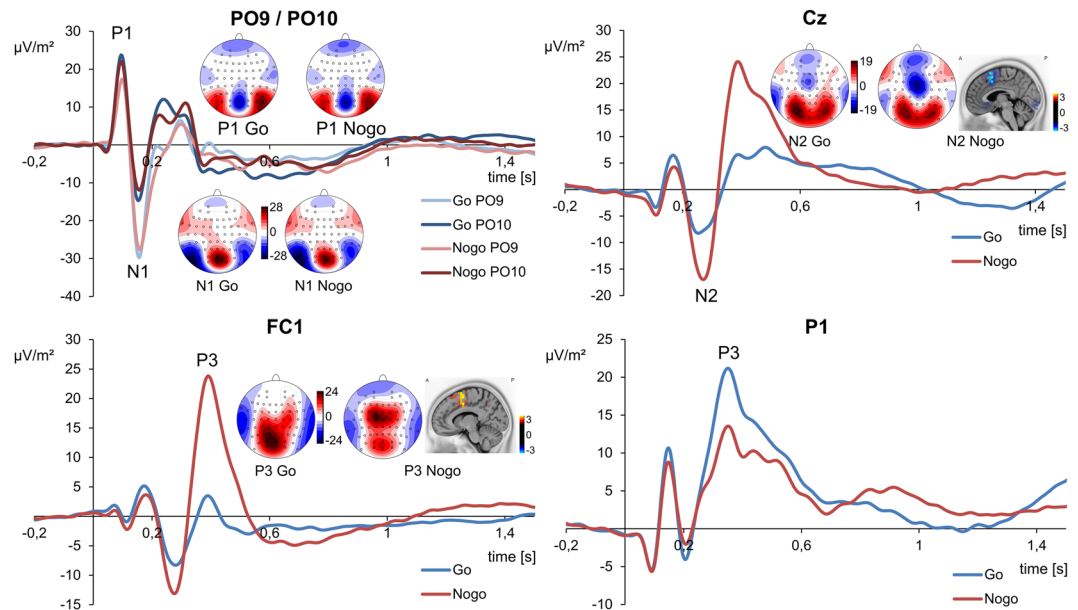
One possible reason for the rather low predictability may relate to the a-priori and literature-based selection of ERP features (as also done in a previous publication of our work group<sup>13</sup>). While the pre-existing literature may help to interpret findings, it might also impede the discovery of novel neurophysiological features, especially “outside” of traditional ERP time windows and / or topographies. Yet, those might still be important to consider when trying to obtain a more comprehensive understanding of neural mechanisms that allow to predict behavioral performance.

To tackle this issue, we chose to investigate motor inhibition (i.e. the capability to interrupt an automated response tendency), as it has already been investigated for decades and is known to reliably produce two different ERPs that closely correlate with behavioral performance, i.e. the frontal-midline Nogo-N2 and the Nogo-P3<sup>2</sup>. While the Nogo-N2 ERP component is assumed to represent pre-motor processes like conflict monitoring, the Nogo-P3 component may reflect the inhibitory process itself or an evaluation process thereof<sup>2,15–18</sup>.

Yet still, ERPs reflect a mixture of different neuronal sources as well as oscillations from various frequency bands<sup>19,20</sup>, but not all of these frequencies are equally important for the investigated cognitive process<sup>21</sup>. This might have important implications for the prediction of behavioral performance. In this context, oscillations in the theta frequency band could be particularly important as they are considered to play a pivotal role in cognitive control and executive functions<sup>22–26</sup>, including the ability to inhibit prepotent motor responses<sup>27,28</sup>. In line with this, it has been shown that response inhibition-related processes are associated with oscillations in the theta frequency band<sup>2,15,29–32</sup>. Yet, the alpha frequency band may also be relevant for inhibitory control<sup>33–37</sup> because oscillations in this frequency band seem to coordinate top-down control processes<sup>34,38</sup>. Furthermore, alpha oscillations may reflect inhibitory mechanisms which control access to task-relevant information<sup>36</sup>, i.e. updating processes of internal task-sets being used<sup>35</sup>. Against this background, it is conceivable that time-frequency (TF) decomposed data may provide a better prediction of response inhibition performance than regular ERP data, as the latter also contains signals from less relevant frequency bands, like beta oscillations. Until now it has however not been directly tested whether theta, beta, or alpha frequency oscillations allow for a better prediction of behavioral response inhibition performance than regular ERPs, which comprise a combination of all of these oscillations. If we found that one of these frequency bands was most predictive for behavioral performance, this would suggest that current views and theoretical conceptions about the importance of the other frequency bands, that are considered to play a role in response inhibition, may need to be refined. Yet, it is also possible that a combination of features (i.e. different frequency bands) may show high predictability of motor response inhibition performance. This would suggest that current views about the frequency architecture underlying motor response inhibition processes, which still often stress the importance of either theta or alpha oscillations, are too simplified.

In the current study, we employed a data-driven feature extraction strategy in combination with support vector machine (SVM) classification approaches to predict inter-individual (i.e. relative, not absolute or categorical) differences in response inhibition performance from objectively identified neurophysiological features, as compared to regular Nogo-N2 and P3 ERPs. The main reason for choosing this approach is that these “classical” inhibition-related ERPs are usually also used to depict inter-individual differences in inhibitory performance. Specifically, we examined which time-frequency features as well as features in the time domain (i.e. ERPs) best predict response inhibition performance, as defined by group membership (good vs. bad performers). For the ERP data, this was combined with source localization methods to examine which functional neuroanatomical structures are associated with these features. Our main reason for using classification instead of regression here is that the brain is a nonlinear system. Consequently, the assumption that there is a linear relation between EEG signal and behavior is not very likely. Non-linear regression methods however require a-priori knowledge about the degree of polynomial function and the nature of the non-linear relationship between the predictor(s) and the dependent variable(s). This is however not (yet) the case in this field of research. Therefore, the primary goal of this work is to find a nexus between neurophysiology and behavior without any prior assumption. Lastly, results from regression analyses are more affected by outliers than SVM-based classification methods. In order to form two equally large groups that differ in performance, we formed a behavioral ratio that accounts for the relation of speed and accuracy by dividing the number of correct inhibitions in the Nogo condition by the response time in Go trials. The reasoning behind this was that previous analyses showed that individuals with low false alarm often respond more slowly to Go trials than individuals with high false alarm rates<sup>13</sup>. Given that slow responses leave more time to carry out a motor inhibition<sup>39</sup>, not accounting for the speed-accuracy-tradeoff would have confounded our analyses. As better performance may hence be reflected by better accuracy (i.e. larger accuracy values) and/or faster responses (i.e. smaller response time values), the ratio we used becomes larger as behavioral performance improves. Based on this, participants with relatively small ratio values may be labeled “bad performers” while participants with relatively large values may be labeled “good performers”. We hence divided the sample into an equal number of good and bad performers (please see methods section for details).

In addition to providing a rigorous evaluation of neurophysiological processes that are important for response inhibition, our data-driven approach also confers another advantage, because it may help to generate new hypotheses<sup>40</sup>. It is therefore possible to derive new, testable hypotheses about the neurophysiological processes and associated functional neuroanatomical structures that determine response inhibition performance. If we found features that complement the currently investigated Nogo-N2 and P3, this would extend or challenge current thinking about the mechanisms underlying the behavioral outcome of response inhibition.



**Figure 1.** Event-related potential (ERP) components on Go and Nogo stimuli presentations. Plots are given for electrodes PO9/PO10 depicting P1 and N1 ERP, electrode Cz for N2, electrode FC1 for Nogo-P3 as well as electrode P1 for Go-P3. ERPs on Go stimuli are shown in blue, ERPs on Nogo stimuli are shown in red. The scalp topography plots reveal the distribution of voltages at the time point of the peak maximum of each ERP component. Time point zero denotes the time point of stimulus delivery. For N2 and P3 ERP, the results of sLORETA source estimation are shown.

## Results

**Behavioral data and group differences.** For the descriptive statistics, the mean and standard error of the mean are given. On average, participants responded correctly in 98.75% ( $\pm 0.10$ ) of Go trials (hits). The average reaction time (RT) was 348 ms ( $\pm 2.73$ ). About 12.12% ( $\pm 0.69$ ) of Nogo trials were not correctly inhibited (false alarms). The mean quotient used for the estimation of behavioral performance (correct inhibition rate divided by mean hit RT) was 0.254 ( $\pm 0.002$ ; Supplementary Fig. 3). Importantly, this transformation accounts for the speed-accuracy tradeoff: good performance is characterized by relatively lower values (due to lower false alarm rates and/or faster Go responses) while poor or “bad” performance is characterized by relatively larger values (due to higher false alarm rates and/or slower Go responses). We used this criterion to form two equally large groups (“good” vs. “bad” performers) using a median split (median = 0.257). Despite the fact that this criterion is relative, rather than absolute or categorical (like the distinction between false and correct trials), it is the most appropriate criterion for our research question, i.e. the mechanisms underlying inter-individual performance variations. The reason for this is that the level/quality of an individual’s task performance can only be assessed in relative comparison to others’ performance the same task.

As evidenced by a Mann-Whitney-U test, the two groups significantly differed with respect to their behavioral ratio values ( $p < 0.001$ ). On average, the low performance group had a ratio of  $0.2313 \pm 0.0018$  while the high performance group had a mean behavioral ratio of  $0.2757 \pm 0.0012$ . The groups furthermore differed with respect to Go hit RTs (low =  $365.1 \text{ ms} \pm 3.56$ ; high =  $331.9 \text{ ms} \pm 0.49$ ) and Nogo FAs (low =  $15.74\% \pm 0.89$ ; high =  $8.56\% \pm 1.91$ ), when tested separately (both  $p < 0.001$ ). The two performance groups did however not differ with respect to sex ( $\chi^2 = 0.817$ ;  $p = 0.366$ ), age ( $p = 0.283$ ), handedness as assessed with the Edinburgh Handedness Inventory ( $p = 0.432$ )<sup>41</sup>, or depression symptoms as assessed with Beck’s depression inventory ( $p = 0.959$ )<sup>42</sup>.

**Standard ERP data analysis.** The P1, N1, N2 and P3 ERP-components in Go and Nogo trials are shown in Fig. 1. For a figure of ERP-components by group, please refer to the Supplementary Material (Supplementary Fig. 1). In any of the analyses described below, the factors handedness and depressive symptoms did not change the pattern of results (all  $F < 0.67$ ;  $p > 0.510$ ).

For the P1 ERP amplitude, the repeated-measures ANOVA revealed a significant main effect of condition ( $F(1, 238) = 10.73$ ;  $p = 0.001$ ;  $\eta^2_p = 0.04$ ). The P1 ERP amplitudes were larger in Go trials ( $19.45 \mu\text{V}/\text{m}^2 \pm 0.74$ ) than in Nogo trials ( $18.67 \mu\text{V}/\text{m}^2 \pm 0.75$ ). The significant main effect of electrodes ( $F(1, 239) = 13.46$ ;  $p < 0.001$ ;  $\eta^2_p = 0.05$ ) suggested that the P1 amplitude was larger at electrode PO10 ( $21.66 \mu\text{V}/\text{m}^2 \pm 1.14$ ) as compared to electrode PO9 ( $16.47 \mu\text{V}/\text{m}^2 \pm 0.88$ ). Additionally, there was a significant interaction effect of condition x electrodes ( $F(1, 238) = 9.22$ ;  $p = 0.003$ ;  $\eta^2_p = 0.04$ ). As shown by a post-hoc t-test ( $t(239) = -3.04$ ;  $p = 0.003$ ), the P1 amplitude difference between Go and Nogo trials was stronger for electrode PO10 ( $1.63 \mu\text{V}/\text{m}^2 \pm 0.36$ ) than for PO9 ( $-0.06 \mu\text{V}/\text{m}^2 \pm 0.38$ ). All other effects including the main effect of group were not significant (all  $F < 0.67$ ;  $p > 0.420$ ).

For the N1 amplitudes, a significant main effect of condition ( $F(1, 238) = 21.43$ ;  $p < 0.001$ ;  $\eta^2 = 0.08$ ) showed that the N1 amplitude was more negative in Nogo trials ( $-12.62 \mu\text{V}/\text{m}^2 \pm 0.95$ ) than in Go trials ( $-10.82 \mu\text{V}/\text{m}^2 \pm 1.00$ ). The significant main effect of electrodes ( $F(1, 238) = 86.65$ ;  $p < 0.001$ ;  $\eta^2_p = 0.27$ ) indicated that N1 amplitudes were more negative at PO9 ( $-18.15 \mu\text{V}/\text{m}^2 \pm 1.20$ ) than at PO10 ( $-5.29 \mu\text{V}/\text{m}^2 \pm 1.16$ ). The interaction effect of condition x electrodes was significant ( $F(1, 238) = 18.20$ ;  $p < 0.001$ ;  $\eta^2_p = 0.07$ ). Subsequent *t*-test ( $t(239) = -4.27$ ;  $p < 0.001$ ) showed that the N1 amplitude difference of Go and Nogo trials was larger for electrode PO9 ( $3.50 \pm 0.61$ ) than for electrode PO10 ( $1.00 \pm 0.50$ ). Additionally, there was a significant interaction effect of condition x group ( $F(1, 238) = 7.27$ ;  $p = 0.008$ ;  $\eta^2_p = 0.03$ ). As shown by a significant post-hoc *t*-test ( $t(210,36) = -2.70$ ;  $p = 0.008$ ), the N1 amplitude difference between Go and Nogo trials was larger in the high performance group ( $5.70 \pm 1.28$ ) than in the low performance group ( $1.50 \pm 0.88$ ). The main effect of group as well as all other interaction effects were not significant (all  $F < 1.34$ ;  $p > 0.249$ ).

For the N2 amplitudes, the ANOVA showed a significant main effect of condition ( $F(1, 238) = 147.76$ ;  $p < 0.001$ ;  $\eta^2_p = 0.38$ ), indicating that the N2 was more negative in Nogo trials ( $-16.22 \mu\text{V}/\text{m}^2 \pm 0.98$ ) than in Go trials ( $-7.91 \pm 0.80$ ). The sLORETA analysis showed that these condition differences in Nogo-N2 amplitudes were associated with activation modulations in the medial frontal cortex and the superior frontal gyrus (BA8) in particular. The main effect of group as well as the interaction effect were not significant (all  $F < 1.97$ ;  $p > 0.162$ ).

For the P3 amplitudes at frontal-central electrode FC1 (i.e. the Nogo-P3), a significant main effect of condition ( $F(1, 238) = 749.51$ ;  $p < 0.001$ ;  $\eta^2_p = 0.76$ ) was found. This indicated that the frontal-central P3 amplitude was larger in Nogo trials ( $23.11 \mu\text{V}/\text{m}^2 \pm 1.04$ ), as compared to Go trials ( $3.27 \mu\text{V}/\text{m}^2 \pm 0.73$ ). The sLORETA analysis showed that these condition differences in P3 amplitudes were associated with modulation of activity in the medial frontal cortex including the medial frontal gyrus (BA6) and the anterior cingulate cortex (ACC, BA24). As indicated by a significant main effect of group ( $F(1, 238) = 10.59$ ;  $p = 0.001$ ;  $\eta^2_p = 0.04$ ), the frontal-central P3 amplitudes were larger in the high performance group ( $15.86 \pm 1.16$ ) than in the low performance group ( $10.51 \pm 1.16$ ). Additionally, there was a significant interaction effect condition x group ( $F(1, 238) = 10.62$ ;  $p = 0.001$ ;  $\eta^2_p = 0.04$ ). To further analyze this interaction effect, the difference of frontal-central P3 amplitudes of Nogo and Go was calculated. A significant post-hoc *t*-test ( $t(238) = 3.26$ ;  $p = 0.002$ ) showed that the frontal-central P3 amplitude difference was larger in the high performance group ( $22.20 \pm 0.94$ ) than in the low performance group ( $17.47 \pm 1.10$ ). For the P3 amplitudes at posterior electrode P1 (i.e. the Go-P3), the ANOVA revealed a significant main effect of condition ( $F(1, 238) = 197.76$ ;  $p < 0.001$ ;  $\eta^2_p = 0.45$ ). This indicated that posterior P3 amplitudes were larger in Go trials ( $17.47 \pm 0.68$ ) in comparison to Nogo trials ( $10.47 \pm 0.70$ ). There was also a main effect of group ( $F(1, 238) = 5.88$ ;  $p = 0.016$ ;  $\eta^2_p = 0.02$ ). Posterior P3 amplitudes were larger in the high performance group ( $15.52 \pm 0.91$ ) than in the low performance group ( $12.42 \pm 0.91$ ). Again, there was a significant interaction effect condition x group ( $F(1, 238) = 5.34$ ;  $p = 0.02$ ;  $\eta^2_p = 0.02$ ). For further analysis of this interaction effect the difference of Nogo and Go P3 amplitudes trials was computed. A significant post-hoc *t*-test ( $t(238) = -2.31$ ;  $p = 0.044$ ) showed that the difference of posterior P3 amplitudes between Nogo and Go trials was larger for the high performance group ( $-8.15 \pm 0.75$ ) than for the low performance group ( $-5.85 \pm 0.65$ ).

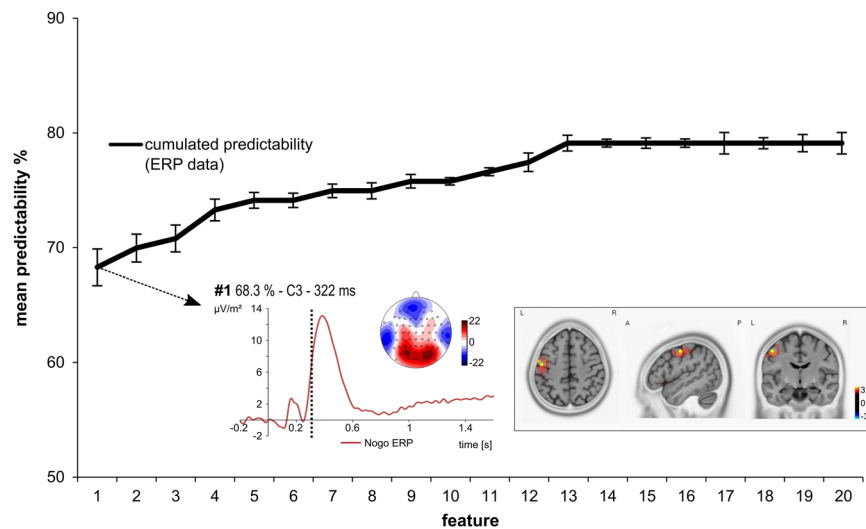
Taken together, the analysis of standard ERP-components revealed the usual pattern of results and that the topography of the Nogo-N2 and Nogo-P3 was centered around fronto-central electrodes<sup>2</sup>.

**Time frequency analysis.** As can be seen in Supplemental Fig. 2 for Nogo trials (right), a power increase is evident in the theta frequency band about 200 to 350 ms post stimulus with a fronto-central topography. To analyze whether this power increase differed between conditions as well as between groups, a repeated measures ANOVA (condition x groups) was calculated. Therefore, the mean theta oscillatory power was determined in the frequency range of 5 to 6 Hz, in the time window of 240 to 320 ms post stimulus for pooled electrodes FCz and Cz. There was a significant main effect of condition ( $F(1,238) = 192.51$ ;  $p < 0.001$ ;  $\eta^2_p = 0.45$ ). Theta power was larger for Nogo trials ( $87.35 \pm 4.70$ ) than for Go trials ( $33.00 \pm 1.55$ ). All other effects were not significant (all  $F < 2.84$ ,  $p > 0.093$ ).

**Correlation analyses.** Linear regression analyses were performed for the above-mentioned ERP components to examine correlations between ERP data and the behavioral data. This was done for each group separately and also including the factor “group” in the regression model (inclusion method). There were no linear correlations between any ERP parameter and the behavioral data (all  $r < 0.2$ ;  $p > 0.3$ ; all  $F < 0.62$ ;  $p > 0.4$ ).

**Machine learning analysis.** As outlined in the methods section we used the *k*-fold method ( $k = 10$ ) to evaluate the predictability of behavioral performance (as assessed via membership in either the ‘good performer’ or ‘bad performer’ group) using ERP and TF-decomposed data. This means that for each extracted feature, there were 10 varying estimations of the predictability of behavioral performance. Using the data from the  $k = 10$  estimations we calculated the 99% confidence bounds for each feature. A significant difference is indicated by no overlap between the calculated 99% confidence bounds. The results of the analysis using ERP data are shown in Fig. 2, where the error bars represent the 99% confidence bounds. Table 1 provides a detailed summary of the ERP features selected by the feature extraction approach. In this context, it must however be understood that while the selected features are confined to a single data/time point, electrode (and frequency in case of the TF-decomposed data), this does not necessarily mean that all other features are not predictive at all. Given that the EEG records a continuous physiological signal showing autocorrelations, it is indeed quite likely that data points, that are neighboring a predictive feature in terms of temporal, topographic, or frequency properties, are also predictive to a certain, albeit lesser extent.

The first ERP-feature (322 ms after Nogo stimulus presentation at electrode C3) led to a prediction accuracy of ~68%, which is significantly different from chance level as indicated by the 99% confidence bounds. Adding more features led to a numerical increase in prediction accuracy (refer Fig. 2 and Table 1), but this increase was



**Figure 2.** Results from the classification analysis using the time domain ERP data. The mean predictability is given depending on the number of features. The black curve in the figure shows the cumulative mean predictability. The error bars represent the 99% confidence level bounds. As can be seen the confidence bounds were overlapping for the first and the second feature. The ERP-curve of the first feature is also shown. The dashed vertical line in the plot shows the time point (feature) at electrode C3. The scalp topography plot reveals the distribution of voltages for this feature. The sLORETA plots (corrected for multiple comparisons using SnPM) show the source of the signal at the time point of the feature.

Feature number	Electrode	Time point (ms)	Mean predictability	p-value
1	C3	322	68%	0.008
2	Cz	1102	70%	0.006
3	CP2	35	71%	0.014
4	CP3	66	73%	0.015
5	TP7	457	74%	0.013
6	CP2	320	74%	0.008
7	C1	1117	75%	0.01
8	CP3	55	75%	0.01
9	CP5	887	76%	0.006
10	CP2	914	76%	0.006
11	C1	301	77%	0.006
12	Cz	1105	78%	0.004
13	C2	74	79%	0.003
14	CP5	406	79%	0.003
15	CP1	320	79%	0.003
16	FC2	270	79%	0.003
17	FC4	367	79%	0.003
18	FC2	1262	79%	0.003
19	F7	98	79%	0.003
20	C2	70	79%	0.003

**Table 1.** Summary of the extracted features showing feature number, electrode site, time point in ms of the extracted feature after stimulus presentation, the mean predictability and the significance as provided from the t-tests used as a filter method in the feature selection step.

not statistically significant, since the 99% confidence bounds largely overlapped with the prediction accuracy obtained after adding the second feature. The sLORETA analysis shows that activation at the time point of this most predictive feature was located in the left motor cortex (M1) (precentral gyrus, BA4) (refer Fig. 2).

As described in the Materials and Methods section, we conducted an add-on analysis with separate training and validation sets in order to evaluate the generalization ability of the selected features. Table 2 provides information on the ERP features selected by the second analysis. Given that the accuracy of selected feature in the

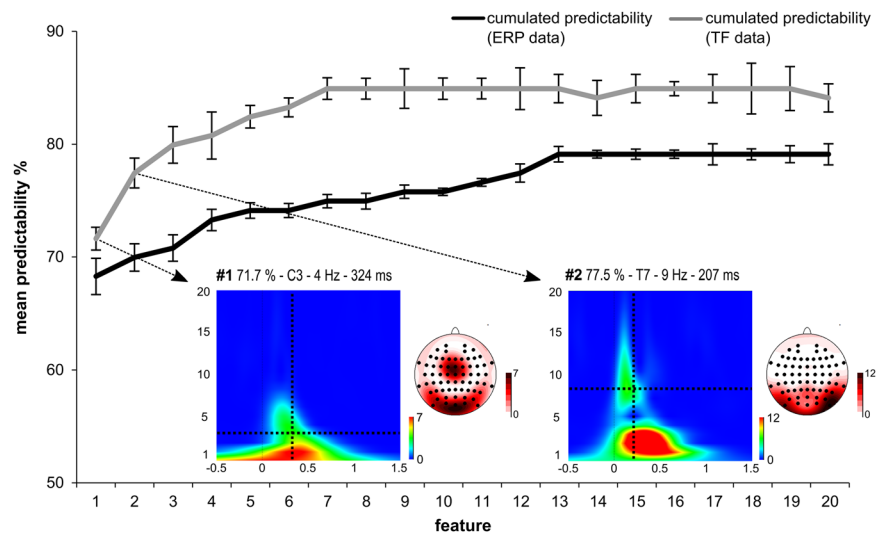


Number of features	Accuracy in training set	Accuracy in validation set	% in which the prediction is better than randomly assigned labels	Accuracy without selected features
1	72%	64%	99.9%	50%
2	76%	65%	99.8%	50%
3	79%	65%	99.7%	50%
4	79%	67%	99.9%	50%
5	80%	67%	100%	50%
6	81%	69%	100%	50%
7	82%	69%	99.9%	50%
8	83%	69%	100%	50%
9	83%	69%	100%	50%
10	83%	68%	100%	50%
11	83%	65%	99.9%	50%
12	83%	67%	99.9%	50%
13	83%	65%	99.8%	50%
14	83%	68%	99.9%	50%
15	83%	69%	99.8%	50%
16	83%	69%	100%	50%
17	83%	69%	100%	50%
18	84%	69%	100%	50%
19	83%	68%	99.9%	50%
20	84%	68%	100%	50%

**Table 2.** Summary of the extracted features for the second analysis with training and validation sets for ERP data. All accuracy values are cumulative.

validation set ranges from 64% up to 69% (see third column in Table 2), we can assume that only the first feature is suitable for prediction. The best predictive feature that was selected in our ERP training set was at 340 ms at electrode C1. Importantly, the best predictive ERP features were extremely similar in both analysis, as they were detected (almost) in the same location (C3 and C1, which are directly neighboring electrodes) and time range (322 ms and 340 ms). Besides that, the accuracy for the best feature in the validation set for ERP data was 64% (i.e. it is above chance level and similar to what we observed without a separate validation set). Without this feature, the prediction accuracy drops to chance level (i.e. 50%), which further indicates the importance of this feature for the classification process. Finally, in 99% of our permutation tests, the feature's prediction accuracy of the real performance group (64%) was higher than the accuracy for randomly assign labels. This means that this feature was not selected by chance. As none of the classical ERP component peaks (such as N2 and P3) were selected by the algorithms, we additionally set out to assess whether and how much selected features are better in predicting inter-individual performance variation/group membership than the classical ERP components. We therefore compared the predictive value of these classical ERP components (as quantified for the classical ERP analysis) to that of the selected feature. To this end, the mean amplitudes of N2 and P3 (quantified in the time window from 250 ms to 280 ms at electrode Cz and from 370 ms to 410 ms at electrodes FC1 and P1) were entered into SVM in order to predict group membership. Of note, the prediction accuracy for these features was 46% and hence at chance level. This clearly demonstrates that conventional ERP components are not always the best features, or even useful for the prediction of behavioral performance. The reason why 46% is at chance level is as follows: There are  $N = 240$  subjects available for the SVM analysis. We used the  $k$ -fold method to validate classifier performance. Since  $n = 10$  this means that there are  $N = 24$  subjects per cross-validation step. Consequently, the resolution for accuracy is 4.17% ( $1/24$ ). Therefore, 46% is within the chance level range. To further evaluate whether this result is biased by the choice of the behavioral performance index, we performed another analysis in which classical ERP components were used as potential features for classifying good and bad performance on the basis of RT data and percentage of correct inhibition in the Nogo condition. Classification accuracy RT was 51% and 55% for the percentage of correct inhibitions in Nogo trials. This did not differ from chance level.

While the results on the basis of ERPs already show that behavioral performance can be predicted with an accuracy of up to ~68% with the help of our automatically selected feature, the results of the analysis using the TF-decomposed data show that prediction accuracy increased to ~78%. The results of the analysis using TF data in comparison to the ERP data are shown in Fig. 3. In Fig. 3 the error bars represent the 99% confidence bounds. Table 3 provides a detailed summary of the TF-features selected by the feature extraction approach. There were two TF-features which significantly increased predictability of behavioral performance. The first TF-feature (324 ms after Nogo stimulus presentation at electrode C3) revealed a predictability of behavioral performance of ~72% (refer Fig. 3 and Table 3) and was found in the theta frequency band (i.e. at 4 Hz). The 99% confidence bounds did not overlap between the first ERP-feature and the first TF-feature indicating that predictability of behavioral performance was significantly better using TF than ERP data, even though both were found almost at the same time point and at the same electrode. The same was the case when adding the second TF-feature (207 ms after Nogo stimulus presentation at electrode T7). The second TF-feature increased predictability to ~78% and



**Figure 3.** Results from the classification analysis using the TF-decomposed data (grey line). The mean predictability is given depending on the number of features. The grey curve in the figure shows the cumulative mean predictability. The error bars represent the 99% confidence level bounds. The black curve showing the cumulative predictability using ERP data is given for comparison. As can be seen the confidence bounds were not overlapping for the first and the second feature (grey curve). The TF plots of the first and the second feature are also shown. The dashed vertical lines in these plots show the time point/frequency (feature) at the respective electrode site (C3 for the first feature and T7 for the second feature). The scalp topography plots reveal the distribution of voltages for these features.

Feature number	Electrode	Hz	time point (ms)	mean predictability	p-value
1	C3	4	324	72%	0.002
2	T7	9	207	78%	0.007
3	F1	7	109	80%	0.006
4	CP1	14	1078	81%	0.003
5	FC1	3	90	83%	0.006
6	CP1	15	1074	83%	0.006
7	C3	12	1023	85%	0.006
8	F4	6	90	85%	0.006
9	Fz	7	20	85%	0.005
10	F4	8	277	85%	0.004
11	FC3	13	684	85%	0.004
12	F4	6	94	85%	0.003
13	POz	14	63	85%	0.005
14	T8	8	121	84%	0.003
15	TP9	10	813	85%	0.004
16	CP3	13	996	85%	0.007
17	T8	7	125	85%	0.001
18	CPz	4	996	85%	0.005
19	C3	11	1023	85%	0.009
20	FC4	7	195	84%	0.001

**Table 3.** Summary of the extracted features showing feature number, electrode site, frequency, time point in ms of the extracted feature after stimulus presentation, the mean predictability and the significance as provided from the t-tests used as a filter method in the feature selection step.

was found in the alpha frequency bands (i.e. at 9 Hz). As can be seen in Fig. 3, the addition of further features numerically increased predictability, but this was not significant given the overlap of the 99% confidence bounds for each of the added features. Importantly, the predictability of the identified features was a little higher for the high performance group (72.45% for the ERP feature alone and 81.67% for the combination of ERP features and the two TF features) than for the low performance group (64.15% for the ERP feature alone and 73.30% for the

Number of features	Accuracy in training set	Accuracy in validation set	% in which the prediction is better than randomly assigned labels	Accuracy without selected features
1	75%	67%	98.3%	50%
2	81%	70%	99.9%	50%
3	83%	70%	97.7%	50%
4	84%	70%	95.2%	50%
5	86%	70%	95.7%	50%
6	86%	70%	99.7%	50%
7	88%	69%	98.8%	50%
8	89%	69%	98.7%	50%
9	90%	69%	99.9%	50%
10	90%	69%	99.7%	50%
11	90%	69%	98.8%	50%
12	90%	70%	96.2%	50%
13	90%	69%	98.3%	50%
14	90%	70%	99.8 5	50%
15	90%	70%	100%	50%
16	90%	70%	99.9%	50%
17	90%	70%	99.7%	50%
18	90%	70%	100%	50%
19	90%	70%	100%	50%
20	90%	70%	100%	50%

**Table 4.** Summary of the extracted features for second analysis with training and validation sets for time frequency data showing feature number, accuracy in training set, accuracy in validation set, corresponding p-value for permutation test and accuracy without selected features. All accuracy values are cumulative.

combination of ERP features and the two TF features). Finally, it is essential to directly compare the classification results obtained using the ERP features and the TF features. Please refer to Tables 1 and 3. For the ERP data it is shown that the best accuracy than we can achieve is about 79% percent with 13 features. However, for TF analysis the highest accuracy is 85% and is achieved with 7 features. However, it needs to be stressed that including more than the best predictive ERP feature, and more than the first two (best predictive) TF-feature did not lead to a statistically significant increase in the classification accuracy. There, above-mentioned maximal prediction accuracy have to be treated with caution. If we just focus on the best feature for each analysis, the TF feature has a better classification accuracy (i.e. 72%) than the ERP-feature (i.e. 68%) (as indicated by non-overlapping confidence bounds).

As described in Materials and Methods section, we conducted an add-on analysis with separate training and validation sets in order to evaluate the generalization ability of the selected features. Table 4 provides information on the TF features selected by the second analysis. For TF data as described in the Table 4, the accuracy in the validation set ranges from 67% to 70%. We therefore only focused on the first feature. The best predictive feature occurred at 340 ms at electrode C3 in 4 Hz. Notably, the best features for TF data are very similar in both analyses (i.e. 324 ms and 340 ms at electrode C3 in 4 Hz). Moreover, prediction accuracy drops to 50% without this feature and in 98% of the permutation tests, the feature's prediction accuracy was higher for the real performance groups than for randomly assigned labels.

## Discussion

In the current study, we used a Go/Nogo task and a machine learning approach to predict inter-individual behavioral performance differences in response inhibition on the basis of neurophysiological data. We used a purely data-driven strategy to extract neurophysiological features that predict response inhibition performance, as assessed by performance group membership. In this context, it should be noted that even though our group formation criterion was relative (rather than absolute or categorical) it is the most appropriate criterion for our research question, i.e. the mechanisms underlying inter-individual performance variations. This focus was set to provide a rigorous test about the relevance of different neurophysiological features and associated functional neuroanatomical structures that are currently considered to best reflect inter-individual variation in inhibitory control. Importantly, we used both time-domain (ERP) and time-frequency decomposed ERP data from a sample of  $N = 240$  subjects in an SVM approach as not all frequencies might be equally important for response inhibition. Additionally, such a data-driven procedure may allow to generate new testable hypotheses<sup>40</sup> about the neurophysiological and functional neuroanatomical processes driving response inhibition processes.

An analysis of the standard ERP-components revealed the usual pattern of results<sup>2</sup>. The results from the SVM analysis showed that both time-domain data (ERPs) and time-frequency (TF) decomposed data collected in Nogo trials allowed to predict behavioral performance above chance level. Yet, predictability was significantly higher using TF-decomposed data (~78%) than when using ERP data (~68%). We had anticipated this effect as ERPs are a mixture of various oscillations, not all of which are (equally) important for the investigated cognitive process (Hoffmann *et al.*<sup>21</sup>).



For the ERP data, a single feature was shown to predict behavioral performance (i.e. group membership) with up to ~68%. While the Nogo-N2 and Nogo-P3 peaks usually constitute the neurophysiological processes that are usually focused on in inhibition research (for review<sup>2</sup>), the predictive feature was identified in-between these two ERPs, i.e. in the transition between these two processes. Moreover, the neuroanatomical source of the feature differed from that of the two ERPs: At the time point of the predictive feature, the sLORETA analysis suggested that left hemisphere motor areas (precentral gyrus, BA4) were most strongly activated (refer Fig. 2). In line with previous work<sup>2,43–46</sup>, the sLORETA analysis however revealed medial frontal areas including the superior frontal gyrus (BA6) and regions in the anterior cingulate cortex (ACC, BA24) to be most strongly activated at the time point of the Nogo-N2 and P3 (refer Fig. 1). Interestingly, the frontal-midline Nogo-N2 component has been suggested to reflect pre-motor processes like conflict monitoring or updating of the response program. In this context, the finding that the predictive feature occurred after the Nogo-N2 and was associated with an activation of the motor cortex (precentral gyrus) seems reasonable and allows for the interpretation that the predictive ERP feature reflects mechanisms where pre-motor processes are transferred to motor programs in order to inhibit the response. This interpretation is corroborated by the finding that responses had to be executed with the right hand / left motor cortex and that the time point of the predictive feature (~322 ms) was before the mean RT (~350 ms). fMRI studies have suggested that the pre-supplementary motor area and the motor cortex are involved in inhibitory control<sup>27,47</sup>. Furthermore, both of these structures show functional interconnectivity via the basal ganglia circuitry that mediates response inhibition. As mentioned, the predictive feature occurred before the Nogo-P3 peak (~405 ms), which has been suggested to reflect evaluative processing of the successful outcome of the inhibition<sup>2,15–17,48,49</sup> and is thus thought to reflect the cognitive process of assessing whether (not) responding was correct<sup>2,50,51</sup>. This interpretation was mainly based on the assumption that the Nogo-P3 peaks too late (i.e. after the mean RT) to reflect motor inhibition processes<sup>52,53</sup> and has been criticized in the past<sup>18</sup>. The fact that the predictive feature was found well before the Nogo-P3 and seems to be related to the motor process however supports the interpretation of evaluative processes being reflected in the Nogo-P3. Even though the Nogo-N2 and Nogo-P3 reflect sub-processes involved in response inhibition<sup>2</sup>, machine learning analysis using ERP data that they are not necessarily important to consider when trying to predict the behavioral outcome of response inhibition, as only the time window in between both ERPs could predict performance. This was furthermore underpinned by the finding that neither the N2, nor the P3 amplitudes could predict performance (i.e. classify group membership) above chance level. While this does not necessarily refute the theory that the Nogo-N2-P3 complex reflects sub-processes of inhibitory control, the results suggest that processes which take part in the transition between both ERPs, should receive much more attention.

Interestingly, the machine learning results obtained for the time-frequency decomposed data revealed a feature at the same electrode (C3) and almost at the same time point (i.e. 324 ms) within the theta frequency band. The convergence of results (i.e. time point and electrode) between ERP and TF-decomposed data further underlines the reliability of the obtained data. In this context, it however needs to be noted that while our predictors are single time points, they are not necessarily isolated events which solely predict group membership. Instead, they are an integral part of cognitive processes, which should be interpreted as a whole for a better understanding of the mechanisms driving response inhibition. Several lines of evidence have already shown that response inhibition processes reflected by the Nogo-N2-P3 complex are associated with oscillations in the theta frequency band<sup>2,15,29–32</sup>. It is therefore plausible that a feature in the theta frequency band was predictive of behavioral performance. This finding is also well in line with literature suggesting that oscillations in the theta frequency band are particularly important for cognitive control and executive functions<sup>22–26</sup>. But while the results showed that the theta frequency feature was most predictive, predictability significantly increased from ~72% to ~78% when additionally considering a second feature in the alpha frequency band ~207 ms after Nogo stimulus presentation. The alpha frequency band has been suggested to be important for inhibitory control<sup>33,36,37</sup>, as oscillations in this frequency band coordinate top-down control processes<sup>34,38</sup>. Specifically, alpha oscillations have been suggested to reflect inhibitory mechanisms that control access to task-relevant information,<sup>36</sup> i.e. updating processes of internal task sets being used for other cognitive processes<sup>35</sup>. Since the predictive alpha band feature was detected at ~207 ms, it was evident in the time window of the P2 ERP component. The P2 has been suggested to reflect resource allocation processes<sup>54–56</sup>. Such processes are closely related to mechanisms which are necessary to update internal task sets and hence processes that are considered to be reflected by alpha oscillations. There were no linear correlations between electrophysiological parameters and behavioral performance, which may be due to the fact that regression approaches and correlations can only model linear interdependencies. This, however, is not the case using SVM approaches.

As the results from the machine learning (SVM) approach were obtained using a data-driven procedure, they call for a change in how we currently conceptualize and investigate response inhibition sub-processes: While theta oscillations and related processes reflected by the Nogo-N2-P3 complex may represent mechanisms operating during the process of inhibitory control<sup>2</sup>, neither the Nogo-N2 nor the Nogo-P3 peak are central to predicting behavioral performance. Rather, it seems that motor cortex-driven processes which take part in the transition between the two ERP peaks are most important for inter-individual differences in response inhibition performance. Currently, most studies have only considered the relevance of *either* theta *or* alpha oscillations for inhibitory control processes. The current results however show that the combination of frequencies best predicts inhibitory control performance and further suggest that the updating of internal task sets should probably receive more attention. Given that the Nogo-N2-P3 complex is found in a wide variety of inhibitory tasks, we would deem it likely that the obtained results are not only valid for Go/Nogo paradigms, but also for other paradigms (e.g. stop-signal tasks). In Go/Nogo paradigms, participants are required to inhibit a prepotent response tendency, whereas participants are asked to stop an requested a motor response in the SST. The latter increases the probability of an initial activation of the (Go) motor response process<sup>49</sup>. Even though both paradigms differ in

the exact nature of the inhibitory processes e.g.<sup>57</sup>, it is generally assumed that both paradigms rely on a similar inhibitory control network<sup>58,59</sup> and are associated with similar electrophysiological correlates<sup>49</sup>.

Regarding the generalizability of the current findings, it however needs to be noted that there are many other machine learning algorithms that can be used for feature extraction and prediction. It is hence possible that results might change when other algorithms/models are used. But in the words of George Box, it can be stated that „*Essentially, all models are wrong, but some are useful*“. Against this background, the usefulness of our results remains to be evaluated in future studies. That is, the functional relevance of the identified processes and functional neuroanatomical structures should be investigated in more detail in the future. However, it needs to be noted that there is (i) accumulating evidence reporting an important role of the motor cortex, but not medial prefrontal areas for response inhibition (see above) and (ii) the entire pattern of results using ERP and TF-decomposed data is conclusive. It may seem arbitrary that subjects were classified by splitting the group into “good” vs. “bad” performers, but this is a necessary step to be able to apply the SVM approach. Yet, it does not necessarily describe the range of individual differences in task performance. Thus, the data may more strongly support the idea that these alternative approaches may extract features that are better predictive of relative task performance. The detection of novel neurophysiological features are possible important to consider in future studies, rather than specifically addressing variability in inter-individual differences.

In summary, our results suggest that when investigating response inhibition, we should not only focus on control-associated ERPs/oscillations and associated prefrontal cortical structures commonly considered to the reflect an important response inhibition network<sup>27</sup>. Instead, we should extend our focus to the motor cortex, theta and alpha frequency oscillations, and the transition between the classical Nogo-N2 and Nogo-P3 correlates of inhibitory control. Using brain stimulation techniques (e.g. TMS or tACS), it would be possible to directly test the importance of the neurophysiological processes identified in the current study.

## Materials and Methods

**Participants.** A sample of  $N = 251$  healthy subjects between 20 and 30 years of age (mean age 23.4 SD = 2.94; 99 females) was recruited for the study. None of the participants reported a history of neurological or mental illness.  $N = 11$  participants were excluded from analysis due to poor performance (as defined by a mean subject's Go RT > mean group Go RT + 3 SD, false alarms and/or misses  $\geq 50\%$ ) or noisy EEG signal (visual inspection: when blinking or other artifacts could not be sufficiently removed with the help of an ICA). The data analysis was performed with  $N = 240$  subjects. Of note, the participants/datasets included in this study have also been included in a previous publication<sup>60</sup>. Yet, we used a different behavioral measure for classification that circumvents the speed-accuracy issues we had in our previous publication, applied stricter exclusion criteria based on the EEG quality and used a vastly different methodological approach which did not limit the number of features to a rather small, theory-driven pre-defined set of ERPs<sup>60</sup>, but instead used the full range of available data. The study and all experimental procedures were approved by the ethics committee of the Ruhr Universität Bochum and TU Dresden. The methods were carried out in accordance with these regulations. Written informed consent was obtained from all participants.

**Task and behavioral parameter for group formation.** We used a standard Go/Nogo task<sup>16,31</sup>, in which two stimuli were presented. Upon presentation of the German word “DRÜCK” (English “press”), the subjects were asked to respond as fast as possible. Upon presentation of the German word “STOPP” (English “stop”), the participants were asked to refrain from responding. The stimuli were presented in the middle of a 17inch CRT screen (100 Hz vertical refresh rate). The ratio of Go and Nogo trials was 70:30. This ratio ensures that there is a strong tendency to respond on Nogo trials making it difficult to refrain from responding upon the presentation of a Nogo stimulus. Subjects were asked to respond within 600 ms. If this response deadline was missed, the word “SCHNELLER” (English “faster”) was presented. This further augments the tendency to respond on Nogo trials. Each trial began with the presentation of a fixation cross in the middle of the screen and was terminated by the participant's first response (correct responses in Go trials or false alarms in Nogo trials). In case of missed Go responses or a correct inhibition of responses in Nogo trials, the trial automatically terminated after 2200 ms had elapsed. The inter-trial interval (ITI) was jittered between 1000 and 1300 ms. On Go trials, responses exceeding this deadline were regarded as “missed” trials. In total, 450 trials were presented. As a performance measure in Go/Nogo tasks, the false alarm rate (i.e. responses upon the presentation of Nogo stimuli) is considered to be most important. However, the likelihood to commit a false alarm also depends on the speed of responding, with a higher response speed leading to higher false alarms rates<sup>39</sup>. To account for this, we calculated an index of false alarms in Nogo trials and the mean reaction time (RT) in Go trials; i.e. the percentage of correct inhibition in the Nogo condition (i.e. 100% minus false alarm rate) divided by mean hit RTs in the Go condition. Importantly, better performance may be reflected by better accuracy (i.e. larger accuracy values) and/or faster responses (i.e. smaller response time values). Based on this ratio, we used a median-split to form two equally large groups. Participants yielding relatively small ratio values were labeled “bad performers” while participants yielding relatively large values were labeled “good performers”. This performance index was used in further analyses using machine learning approaches; i.e. it was examined in how far performance group membership was predictable on the basis of neurophysiological features extracted from the neurophysiological data.

**EEG recording and analysis.** The EEG was recorded from 64 Ag/AgCl electrodes using the extended 10/20 system against a reference electrode placed at electrode FCz (QuickAmp, Brain Products Inc.). The signal was automatically re-referenced to a common average reference by the amplifier. The sampling rate was 1 kHz. Electrode impedances were kept below 5 k $\Omega$ . After recording, the data were down-sampled to 256 Hz. The EEG was digitally filtered off-line using IIR band-width filters at 0.5 and 18 Hz (each with a slope of 48 dB/oct) using the BrainVision Analyzer 2 software package (BrainProducts, Inc.). Then, gross artifacts were manually removed

from the EEG. Following this step, horizontal and vertical eye-movements as well as pulse artifacts were detected using independent component analysis (ICA) (infomax algorithm), and components reflecting these artifacts were removed. After reconstructing the EEG from the remaining components, electrode FCz was interpolated using a spherical spline interpolation. Then, the EEG was segmented into Go and Nogo trials. Only correct Go and Nogo trials were included in the data analysis. Only Go trials where the correct response was carried out within 1200 ms after stimulus onset were considered as correct. Nogo trials were included when no response was given in an interval of 2200 ms after stimulus presentation. The segments were 4000 ms long, starting 2000 ms before the locking time point (time point zero) and ended 2000 ms after this locking time point. Though this leads to an overlap between succeeding trials, this long epoch is needed for the time-frequency decomposition step to allow a reliable assessment of the power in low frequency bands<sup>61</sup>. On average, 287.7 ( $\pm 8.8$ ) Go trials and 106.9 ( $\pm 12.9$ ) Nogo trials were included. This segmentation step was followed by an automated artifact rejection procedure to eliminate any artifacts that might have survived the prior data inspection. A maximal value difference above 200  $\mu\text{V}$  in a 100 ms interval as well as an activity below 0.5  $\mu\text{V}$  in a 200 ms period were used as rejection criteria. About 2.7% of the trials were rejected due to artifacts. Next, a current source density (CSD) transformation<sup>62</sup> (order of splines  $m = 4$ , maximum degree of the Legendre polynomials  $n = 10$ , precision of  $2.72^{-7}$ ) was applied to re-reference the data. Due to this, the resulting CSD values are given in  $\mu\text{V}/\text{m}^2$ . Baseline correction was applied in the time range from  $-200$  ms to 0 ms (i.e. prior to target onset) before the segments were separately averaged for Go and Nogo on a single subject level to derive the event-related potential (ERP). All ERP components were quantified using local maximum peak detection on single subject level. For P1, a time range of 90 to 110 ms at electrodes PO9 and PO10 was used. For N1, the time range was set to 170 to 190 ms at electrodes PO9 and PO10. N2 ERP mean amplitudes were determined in the time range from 250 to 280 ms at electrode Cz. For the P3, the mean amplitudes were quantified from 370 to 410 ms at electrodes FC1 and P1. All electrodes were chosen based on scalp topography of each ERP component and condition. This procedure was validated as described by Mückschel *et al.*<sup>63</sup> A time window of maximum activity was identified for each component. Using these time windows, the mean amplitude was extracted for every electrode. Each electrode was then compared against the average of all other electrodes using Bonferroni correction ( $p = 0.0007$ ). Only electrodes with significantly larger amplitudes than the remaining electrodes were selected. In addition to this time-domain analysis of the data, a time-frequency analysis was conducted by means of a continuous wavelet transformation (CWT) employing Morlet wavelets ( $w$ ) in the time domain to different frequencies ( $f$ ):

$$W(t, f) = A \exp(-t^2/\sigma_t^2) \exp(2i\pi f t)$$

where  $t$  = time,  $A = (\sigma_t \sqrt{\pi})^{-1/2}$ ,  $\sigma_t$  = wavelet duration, and  $i = \sqrt{-1}$ . For analysis and TF-plots, a ratio of  $f_0/\sigma_f = 5.5$  was used, where  $\sigma_f$  is the width of the Gaussian shape in the frequency domain and  $f_0$  is the central frequency. The analysis was conducted in the frequency range from 1 to 18 Hz, and a central frequency at 1 Hz intervals was employed. For different  $f_0$ , time and frequency resolutions [or wavelet duration and spectral bandwidth] can be calculated as  $2\sigma_t$  and  $2\sigma_f$  respectively.  $\sigma_t$  and  $\sigma_f$  are related by the equation  $\sigma_t = 1/(2\pi\sigma_f)$ . For example, for  $f_0 = 1$  Hz,  $2\sigma_t = 1770$  ms and  $2\sigma_f = 0.36$  Hz; for  $f_0 = 3$  Hz,  $2\sigma_t = 580$  ms and  $2\sigma_f = 1.09$  Hz; for  $f_0 = 5$  Hz,  $2\sigma_t = 350$  ms and  $2\sigma_f = 1.82$  Hz. The total power was calculated by performing TF decomposition on the single trial level before averaging.

**Data-driven feature extraction procedure and support vector machine (SVM) analysis.** Based on a median split of the behavioral performance parameter (i.e. correct inhibition rate on Nogo trials, relative to the RT on Go trials), two groups of subjects were created: a “high performance group” and a “low performance group”. Thereafter, a machine learning approach was employed to predict group membership on the basis of the neurophysiological data from Nogo trials. Only neurophysiological data from correct Nogo trials, i.e. trials where no response was executed, was used.

For the time-domain features, all time points from zero (i.e. locking time point/time point of stimulus presentation) to 1.5 seconds with the resolution of the sampling frequency (256 Hz) were extracted as possible features. This was done for each of the 64 channels and for every subject. For the time-frequency decomposed data the same approach was applied for every frequency in the frequency range between 1 and 18 Hz using 1 Hz frequency steps. Moreover, before using feature selection and classification methods, and in order to eliminate the effect of different range of features on classification performance all features were normalized into a z-score. This was done for two reasons: First, it increases the convergence speed of feature detection algorithms<sup>64</sup>. Second, features may govern/bias the feature detection algorithm in case they have different value ranges. This problem is circumvented using z-transformation<sup>65</sup>, because z-transformation makes all features have a mean of zero and a standard deviation equal to one. After normalizing the features, the feature selection was applied. This is crucial for machine learning algorithms, since it eliminates surplus/irrelevant features and reduces the problem of having a ‘small’ data set relative to the size of the possible feature set. Both of these factors can otherwise reduce the classifier performance. In the feature selection stage, an optimal subset of features is selected from the original feature set. Feature selection algorithms roughly divide into two categories: “filter” and “wrapper” methods<sup>66</sup>. The filter methods select a subset of features according to general characteristics of the data, independently of chosen classifier. Wrapper methods require a predetermined classifier and evaluate features according to their performance to discriminate between classes<sup>66</sup>. Usually, wrapper methods lead to better results than filter methods, because the selected features are based on classifier performance<sup>67</sup>. However, they can be significantly slower than filter methods. One way to overcome this problem is to combine filter and wrapper methods: i.e. filter methods are applied first to select some features. Then, these selected features are used as input for wrapper methods. This was done in the current study using MATLAB 2017a (Mathworks Inc.). In particular, t-test and sequential floating forward selection (SFFS) methods are employed as a filter and wrapper methods, respectively<sup>67</sup>. For initial sorting of the

features, a t-test is calculated between the two groups created using the median split procedure for each time point (i.e. feature). If the p-value is below 0.01 this time point (feature) is selected (the precise p-values are given in Table 1 and Table 3 in the results section). Then these selected features are used as input for the SFFS algorithm. SFFS combines two separate algorithms;<sup>68,69</sup> sequential forward selection (SFS) and sequential backward selection (SBS). SFS starts from an empty set of features and sequentially adds features that result in the highest classifier accuracy when being combined with the features that have already been selected. SBS works in the opposite direction. In SFFS, each feature selection step comprises SFS and SBF<sup>68,69</sup> and were implemented in MATLAB 2017a (Mathworks Inc.). Following SFFS, selected features are fed to a support vector machine (SVM) employing a radial basis function (RBF) kernel with sigma equals to 5. This was done using MATLAB 2017a (Mathworks Inc.) and the LIBSVM toolbox<sup>70</sup>. SVMs are supervised learning algorithms that aim towards an optimal separation of distinct classes. In other words, the applied classifier projects input data into high-dimensional feature space to determine a hyper-plane which is able to separate the groups.

Importantly, the result of the SVM method was cross-validated in this study. Since neuroimaging studies usually deal with a ‘small’ number of subjects, it is very important to use cross-validation methods. There are two popular methods that can be used: Leave One Out Cross Validation (LOOCV) and k-fold cross-validation<sup>71</sup>. However, the high variance of the classification accuracy and the computation time are two major problems of the LOOCV method<sup>72</sup>. The k-fold method randomly divides the data into k portions in which k-1 portion is considered as training data and other as testing data. By continuing this k-times all subjects in the data set are part of the testing and training set. The resulting classification accuracy is the average of the all k-folds<sup>71</sup>. Usually, the value of k is between 5 to 10 in machine learning. We used k = 10 in this study. This means that for each extracted feature there were 10 estimations of the predictability of behavioral performance. Using the data from the k = 10 estimations we calculated the 99% confidence bounds for each feature. These confidence bounds were then used to examine (i) in how far the different features provided a significant increase predictability of behavioral performance, and (ii) in how far there is a difference in the predictability of behavioral performance between ERP and TF-decomposed data. A significant difference is indicated by no overlap between the calculated 99% confidence bounds. Importantly, the applied machine learning approach also minimizes the risk of ending up with false positive features: While there is indeed still a small risk for false positive features to survive feature selection and enter the machine learning approach, the subsequent k-fold validation procedure, which is an integral part of the machine learning approach, minimized the risk of any false positive being ultimately selected as a predictive feature because it mixes and recombines the sample many times, which strongly decreases the likelihood of false positives having a strong and consistent effect.

Since the feature selection and evaluation steps were done in the same data set, it could be argued that the generalization ability (i.e. working well on other data sets) required further verification. In order to demonstrate the validity of the initially applied k-fold method and verify that the selected features do not have an overfitting problem (i.e. that they do not work well only on the training set), we ran an additional analysis, for which the data were divided into two separate data sets (training and validation). 70% of the subjects were used as the training set (168 subjects) and the remaining subjects (72) as the validation set. In the training dataset, we again employed the k-fold and feature selection methods (t-test and SFFS) in order to achieve maximal comparability. For validation of the results (i.e., the features that had been selected in the first data set), we used SVM in the validation set. In order to have a comprehensive assessment of selected features, we also used permutation tests for the features selected from the training set and applied it on the validation set. For those permutation tests, the data were randomly divided into two groups and SVM was used for the prediction of those random groups. This process was repeated 1000 times. In the end, we calculated the percentage of how many times (out of 1000) the selected feature from the training set predicted group membership better than a randomly assigned group label. To confirm that the selected features are practically important for classification accuracy (i.e. to demonstrate that the accuracy will drop to chance level without them) we investigated the impact of omitting these features on classification accuracy. To this end, we started to consecutively omit the best features and ran the SVM without them. In this context, it is worth mentioning that we trained the classifier on the training set and the results are just based on the validation set.

**Source localization analysis.** For each of the time-domain (ERP) features which were shown to be predictive for behavioral performance in the SVM analysis step (see results section) a source localization analysis was conducted. For this analysis, sLORETA (standardized low resolution brain electromagnetic tomography;<sup>73</sup> was used, which provides a single solution to the inverse problem<sup>73,74</sup>. For sLORETA, the intracerebral volume is partitioned into 6239 voxels at 5 mm spatial resolution. Then, the standardized current density at each voxel is calculated in a realistic head based on the MNI152 template. It has been mathematically proven that sLORETA provides reliable results without a localization bias<sup>75</sup>. Moreover, there is evidence from EEG/fMRI and neuronavigated EEG/TMS studies underlining the validity of the sources estimated using sLORETA<sup>21,76</sup>. The voxel-based sLORETA images were compared against zero using the sLORETA-built-in one sample t-test. Importantly, voxel-wise randomization tests with 2000 permutations, based on statistical nonparametric mapping (SnPM) were applied to control the risk of false positive results. Significant voxels ( $p < 0.01$ , corrected for multiple comparisons) were located in the MNI-brain.

**Statistics.** A standard analysis of the electrophysiological data was performed using repeated measures ANOVAs including the within-subject factors “condition” (Go vs. Nogo trials) and “electrode” as well as the between-subject factor group (low vs high performance) in the model. Greenhouse-Geisser correction was applied wherever appropriate and all post-hoc tests were Bonferroni corrected. As indicated by Shapiro-Wilks tests and confirmed by visual inspection, all analyzed variables were normally distributed (all  $W > 0.51$ ;  $df = 239$ ;  $p > 0.212$ ). For the descriptive statistics, the mean and standard error of the mean are given.



## Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

- Cohen, M. X. *Analyzing neural time series data: theory and practice* (MIT Press 2014).
- Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M. & Herrmann, C. S. Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **87**, 217–233 (2013).
- Keil, A. *et al.* Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* **51**, 1–21 (2014).
- Stock, A.-K., Gohil, K., Huster, R. J. & Beste, C. On the effects of multimodal information integration in multitasking. *Sci. Rep.* **7**, 4927 (2017).
- Nunez, P. L. *et al.* EEG coherency. I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* **103**, 499–515 (1997).
- Huster, R. J., Plis, S. M. & Calhoun, V. D. Group-level component analyses of EEG: validation and evaluation. *Front. Neurosci.* **9**, 254 (2015).
- Brunner, J. F. *et al.* Long-term test-retest reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo wave in a visual Go/NoGo task. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **89**, 106–114 (2013).
- Mückschel, M., Chmielewski, W., Ziemssen, T. & Beste, C. The norepinephrine system shows information-content specific properties during cognitive control - Evidence from EEG and pupillary responses. *NeuroImage* **149**, 44–52 (2017).
- Ouyang, G., Herzmann, G., Zhou, C. & Sommer, W. Residue iteration decomposition (RIDE): A new method to separate ERP components on the basis of latency variability in single trials. *Psychophysiology* **48**, 1631–1647 (2011).
- Bridwell, D. A. *et al.* Moving Beyond ERP Components: A Selective Review of Approaches to Integrate EEG and Behavior. *Front. Hum. Neurosci.* **12** (2018).
- Boehler, C. N. *et al.* Sensory MEG responses predict successful and failed inhibition in a stop-signal task. *Cereb. Cortex N. Y. N* **1991**(19), 134–145 (2009).
- Plewan, T., Wascher, E., Falkenstein, M. & Hoffmann, S. Classifying Response Correctness across Different Task Sets: A Machine Learning Approach. *PloS One* **11**, e0152864 (2016).
- Stock, A.-K., Popescu, F., Neuhaus, A. H. & Beste, C. Single-subject prediction of response inhibition behavior by event-related potentials. *J. Neurophysiol.* **115**, 1252–1262 (2016).
- Combrisson, E. & Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **250**, 126–136 (2015).
- Beste, C., Ness, V., Falkenstein, M. & Saft, C. On the role of fronto-striatal neural synchronization processes for response inhibition—evidence from ERP phase-synchronization analyses in pre-manifest Huntington's disease gene mutation carriers. *Neuropsychologia* **49**, 3484–3493 (2011).
- Beste, C., Willemsen, R., Saft, C. & Falkenstein, M. Response inhibition subprocesses and dopaminergic pathways: basal ganglia disease effects. *Neuropsychologia* **48**, 366–373 (2010).
- Beste, C., Dziobek, I., Hielscher, H., Willemsen, R. & Falkenstein, M. Effects of stimulus-response compatibility on inhibitory processes in Parkinson's disease. *Eur. J. Neurosci.* **29**, 855–860 (2009).
- Wessel, J. R. & Aron, A. R. It's not too late: the onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. *Psychophysiology* **52**, 472–480 (2015).
- Başar, E. *Brain Function and Oscillations: Volume I: Brain Oscillations. Principles and Approaches* (Springer-Verlag 1998).
- Başar, E., Başar-Eroglu, C., Karakaş, S. & Schürmann, M. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *Int. J. Psychophysiol.* **39**, 241–248 (2001).
- Hoffmann, S., Labrenz, F., Themann, M., Wascher, E. & Beste, C. Crosslinking EEG time-frequency decomposition and fMRI in error monitoring. *Brain Struct. Funct.* **219**, 595–605 (2014).
- Cavanagh, J. F., Zambrano-Vazquez, L. & Allen, J. J. B. Theta lingua franca: a common mid-frontal substrate for action monitoring processes. *Psychophysiology* **49**, 220–238 (2012).
- Cavanagh, J. F. & Frank, M. J. Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* **18**, 414–421 (2014).
- Cohen, M. X. A neural microcircuit for cognitive conflict detection and signaling. *Trends Neurosci.* **37**, 480–490 (2014).
- De Blasio, F. M. & Barry, R. J. Prestimulus delta and theta determinants of ERP responses in the Go/NoGo task. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **87**, 279–288 (2013).
- Harper, J., Malone, S. M. & Bernat, E. M. Theta and delta band activity explain N2 and P3 ERP component activity in a go/no-go task. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **125**, 124–132 (2014).
- Bari, A. & Robbins, T. W. Inhibition and impulsivity: behavioral and neural basis of response control. *Prog. Neurobiol.* **108**, 44–79 (2013).
- Diamond, A. Executive functions. *Annu. Rev. Psychol.* **64**, 135–168 (2013).
- Beste, C. *et al.* Striosomal dysfunction affects behavioral adaptation but not impulsivity—Evidence from X-linked dystonia-parkinsonism. *Mov. Disord. Off. J. Mov. Disord. Soc.* <https://doi.org/10.1002/mds.26895> (2017).
- Dippel, G., Chmielewski, W., Mückschel, M. & Beste, C. Response mode-dependent differences in neurofunctional networks during response inhibition: an EEG-beamforming study. *Brain Struct. Funct.* 1–11 <https://doi.org/10.1007/s00429-015-1148-y> (2015).
- Ocklenburg, S., Güntürkün, O. & Beste, C. Lateralized neural mechanisms underlying the modulation of response inhibition processes. *NeuroImage* **55**, 1771–1778 (2011).
- Quetscher, C. *et al.* Striatal GABA-MRS predicts response inhibition performance and its cortical electrophysiological correlates. *Brain Struct. Funct.* **220**, 3555–3564 (2015).
- Doppelmayr, M., Klimesch, W., Hödlmoser, K., Sauseng, P. & Gruber, W. Intelligence related upper alpha desynchronization in a semantic memory task. *Brain Res. Bull.* **66**, 171–177 (2005).
- Klimesch, W.  $\alpha$ -band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **16**, 606–617 (2012).
- Klimesch, W. Evoked alpha and early access to the knowledge system: the P1 inhibition timing hypothesis. *Brain Res.* **1408**, 52–71 (2011).
- Klimesch, W., Sauseng, P. & Hanslmayr, S. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* **53**, 63–88 (2007).
- Sauseng, P. *et al.* The interplay between theta and alpha oscillations in the human electroencephalogram reflects the transfer of information between memory systems. *Neurosci. Lett.* **324**, 121–124 (2002).
- Klimesch, W., Freunberger, R. & Sauseng, P. Oscillatory mechanisms of process binding in memory. *Neurosci. Biobehav. Rev.* **34**, 1002–1014 (2010).
- Verbruggen, F. & Logan, G. D. Models of response inhibition in the stop-signal and stop-change paradigms. *Neurosci. Biobehav. Rev.* **33**, 647–661 (2009).
- Kell, D. B. & Oliver, S. G. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **26**, 99–105 (2004).
- Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).



42. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **4**, 561–571 (1961).
43. Bodmer, B. & Beste, C. On the dependence of response inhibition processes on sensory modality. *Hum. Brain Mapp.* <https://doi.org/10.1002/hbm.23495> (2017).
44. Bokura, H., Yamaguchi, S. & Kobayashi, S. Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **112**, 2224–2232 (2001).
45. Fallgatter, A. J. *et al.* Altered response control and anterior cingulate function in attention-deficit/hyperactivity disorder boys. *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **115**, 973–981 (2004).
46. Lavric, A., Pizzagalli, D. A. & Forstmeier, S. When 'go' and 'nogo' are equally frequent: ERP components and cortical tomography. *Eur. J. Neurosci.* **20**, 2483–2488 (2004).
47. Duann, J.-R., Ide, J. S., Luo, X. & Li, C. R. Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *J. Neurosci. Off. J. Soc. Neurosci.* **29**, 10171–10179 (2009).
48. Beste, C., Stock, A.-K., Epplen, J. T. & Arning, L. Dissociable electrophysiological subprocesses during response inhibition are differentially modulated by dopamine D1 and D2 receptors. *Eur. Neuropsychopharmacol. J. Eur. Coll. Neuropsychopharmacol.* **26**, 1029–1036 (2016).
49. Enriquez-Geppert, S., Konrad, C., Pantev, C. & Huster, R. J. Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage* **51**, 877–887 (2010).
50. Chun, J. *et al.* Can P300 distinguish among schizophrenia, schizoaffective and bipolar I disorders? An ERP study of response inhibition. *Schizophr. Res.* **151**, 175–184 (2013).
51. Hillman, C. H. *et al.* From ERPs to academics. *Dev. Cogn. Neurosci.* **2**(Suppl 1), S90–98 (2012).
52. Roche, R. A. P., Garavan, H., Foxe, J. J. & O'Mara, S. M. Individual differences discriminate event-related potentials but not performance during response inhibition. *Exp. Brain Res.* **160**, 60–70 (2005).
53. Schmajuk, M., Liotti, M., Busse, L. & Woldorff, M. G. Electrophysiological activity underlying inhibitory control processes in normal adults. *Neuropsychologia* **44**, 384–395 (2006).
54. Campbell, J. & Sharma, A. Compensatory changes in cortical resource allocation in adults with hearing loss. *Front. Syst. Neurosci.* **7**, 71 (2013).
55. Geisler, M. W. & Murphy, C. Event-related brain potentials to attended and ignored olfactory and trigeminal stimuli. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **37**, 309–315 (2000).
56. Sugimoto, F. & Katayama, J. Somatosensory P2 reflects resource allocation in a game task: assessment with an irrelevant probe technique using electrical probe stimuli to shoulders. *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.* **87**, 200–204 (2013).
57. Verbruggen, F. & Logan, G. D. Automatic and Controlled Response Inhibition: Associative Learning in the Go/No-Go and Stop-Signal Paradigms. *J. Exp. Psychol. Gen.* **137**, 649–672 (2008).
58. Aron, A. R., Robbins, T. W. & Poldrack, R. A. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn. Sci.* **18**, 177–185 (2014).
59. Rubia, K. *et al.* Mapping Motor Inhibition: Conjunctive Brain Activations across Different Versions of Go/No-Go and Stop Tasks. *NeuroImage* **13**, 250–261 (2001).
60. Stock, A.-K., Popescu, F., Neuhaus, A. H. & Beste, C. Single-subject prediction of response inhibition behavior by event-related potentials. *J. Neurophysiol. Jn.* **00969**, 2015, <https://doi.org/10.1152/jn.00969.2015> (2015).
61. Beste, C. *et al.* Functional 5-HT<sub>1A</sub> receptor polymorphism selectively modulates error-specific subprocesses of performance monitoring. *Hum. Brain Mapp.* **31**, 621–630 (2010).
62. Nunez, P. L. & Pilgreen, K. L. The spline-Laplacian in clinical neurophysiology: a method to improve EEG spatial resolution. *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.* **8**, 397–413 (1991).
63. Mückschel, M., Stock, A.-K. & Beste, C. Psychophysiological Mechanisms of Interindividual Differences in Goal Activation Modes During Action Cascading. *Cereb. Cortex* **24**, 2120–2129 (2014).
64. Theodoridis, S. & Koutroumbas, K. *Pattern recognition* (Elsevier/Acad. Press 2011).
65. Raschka, S. *Python machine learning: unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analysis* (Packt Publishing Limited 2015).
66. Guyon, I. & Elisseeff, A. An Introduction to Variable Feature Selection. *J. Mach. Learn. Res.* 1157–1182 (2003).
67. Saeys, Y., Inza, I. & Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507–2517 (2007).
68. Chandrashekar, G. & Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **40**, 16–28 (2014).
69. Khazaei, A., Ebrahimzadeh, A. & Babajani-Feremi, A. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav.* **10**, 799–817 (2016).
70. Chang, C.-C. & Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans Intell Syst Technol* **2**(27), 1–27: 27 (2011).
71. Arlot, S. & Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010).
72. *Pattern Recognition with Support Vector Machines.* **2388**, (Springer Berlin Heidelberg 2002).
73. Pascual-Marqui, R. D. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* **24**(Suppl D), 5–12 (2002).
74. Marco-Pallarés, J., Grau, C. & Ruffini, G. Combined ICA-LORETA analysis of mismatch negativity. *NeuroImage* **25**, 471–477 (2005).
75. Sekihara, K., Sahani, M. & Nagarajan, S. S. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *NeuroImage* **25**, 1056–1067 (2005).
76. Dippel, G. & Beste, C. A causal role of the right inferior frontal cortex in implementing strategies for multi-component behaviour. *Nat. Commun.* **6**, 6587 (2015).

## Acknowledgements

This work was partly supported by a Grant from the Deutsche Forschungsgemeinschaft (DFG) SFB 940 project B8.

## Author Contributions

M.M., A.K. collected the data, M.M., A.K., C.B., A.V., A.H. performed the statistical analyses of the behavioural data and wrote the manuscript. M.M., A.K., C.B., A.V. analysed the data. A.K. and C.B. designed the task and the E.E.G. analysis and reviewed the manuscript. A.K., M.M. and C.B. conceptualised the study and reviewed the manuscript. M.M., A.K., C.B., A.V. analysed the E.E.G. data.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-34727-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018