Exploratory Report

The P300 as marker of inhibitory control – Fact or fiction?

René J. Huster a,b,*, Mari S. Møller a,b,c, Christina Thunberg a,b and Liisa Raud a,b

a Multimodal Imaging and Cognitive Control Lab, Department of Psychology, University of Oslo, Oslo, Norway
b Cognitive and Translational Neuroscience Cluster, Department of Psychology, University of Oslo, Oslo, Norway
c Sunnaas Rehabilitation Hospital, Nesodden, Norway

Article info

Article history:
Received 19 August 2019
Reviewed 9 September 2019
Revised 31 December 2019
Accepted 11 May 2020
Published online 5 August 2020

Keywords:
P300
P3
N200
N2
Stop signal task
Stopping
Inhibition
Latency

Abstract

Inhibitory control, i.e., the ability to stop or suppress actions, thoughts, or memories, represents a prevalent and popular concept in basic and clinical neuroscience as well as psychology. At the same time, it is notoriously difficult to study as successful inhibition is characterized by the absence of a continuously quantifiable direct behavioral marker. It has been suggested that the P3 latency, and here especially its onset latency, may serve as neurophysiological marker of inhibitory control as it correlates with the stop signal reaction time (SSRT). The SSRT estimates the average stopping latency, which itself is unobservable since no overt response is elicited in successful stop trials, based on differences in the distribution of go reaction times and the delay of the stop-relative to the go-signal in stop trials.

In a meta-analysis and an independent electroencephalography (EEG) experiment, we found that correlations between the P3 latency and the SSRT are indeed replicable, but also unspecific. Not only does the SSRT also correlate with the N2 latency, but both P3 and N2 latency measures show similar or even higher correlations with other behavioral parameters such as the go reaction time or stopping accuracy. The missing specificity of P3–SSRT correlations, together with the general pattern of associations, suggests that these manifest effects are driven by underlying latent processes other than inhibition, such as behavioral adaptations in context of performance monitoring operations.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The ability to quickly stop ongoing actions, thoughts, or memories, is considered a hallmark of executive functions or cognitive control. Impaired inhibitory control has consequently been associated with a number of mental disease states, including attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder, or substance use...
disorders (e.g., Nigg et al., 2007). The study of inhibition in the motor domain, so-called response inhibition, serves as a proxy for more cognitive domains in which the actual effects of interest, such as the suppression of memories or urges, are notoriously hard to observe (Aron, 2007).

Response inhibition in humans is most commonly studied using the stop-signal task (SST), which putatively probes the rapid suppression of an already initiated and predominant response. In short, a go-signal is presented in the majority of trials (e.g., 75%), and participants are instructed to respond as quickly as possible with a button press. In the remaining trials, a stop-signal follows the go-signal with a short delay (the stop-signal delay – SSD) and instructs the participants to withhold the response. By systematically varying the SSD such that participants are unsuccessful at stopping in about 50% of the stop trials, the SST allows for the calculation of the stop-signal reaction time (SSRT; e.g., Band et al., 2003). The SSRT, often calculated by subtracting the average SSD from the mean response time to go stimuli (goRT), provides an estimate of an individual’s speed of the stopping process, and is generally considered the purest parameter representing inhibitory control capabilities (Band et al., 2003; but also see Matzke et al., 2019; Skippen et al., 2019).

A network formed by the right inferior frontal gyrus, the pre-supplementary motor area, and distinct nuclei of the basal ganglia such as the subthalamic nucleus, are believed to implement the stopping process at the neural level (e.g., Aron et al., 2014). While the bulk of research on this inhibitory control network has been done on the stopping of behavioral responses, recent research suggests that it may be domain-general and thus extends its functions also into the more cognitive realm (e.g., Wessel et al., 2016). Much of the work leading to the identification of this network has been done using functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (e.g., Cai et al., 2012). However, since fMRI suffers from a low temporal resolution, recent research has shifted more towards the use of electroencephalography (EEG) to better understand the temporal dynamics of inhibitory control (e.g., Huster et al., 2013). The identification of an unequivocal neural signature of an inhibitory control process is of utmost importance, because the SSRT is prone to strategic as well as maladaptive adjustments of behavior (such as reflected in response slowing or go omissions; e.g., Matzke et al., 2019; Verbruggen et al., 2019; Verbruggen et al., 2013). A direct neural indicator of inhibitory control could thus serve as diagnostic marker in disorders believed to be associated with deficient inhibition. Not least, a valid neural fingerprint of inhibition proper would constitute a meaningful target for neuromodulatory interventions aiming at the augmentation of inhibitory control capabilities.

One event-related potential (ERP) that has repeatedly been suggested as a potential marker of inhibition is the fronto-central P300, or P3a (e.g., Enriquez-Geppert et al., 2010; Huster et al., 2013; Wessel & Aron, 2015). This positive potential, henceforth simply referred to as P3, is consistently found following stimuli that instruct participants not to respond to a stimulus in a context where responding is the prepotent tendency. Increased fronto-central activity in the low theta and delta frequency range with a relatively strong time- or phase-locking drive the P3 ERP (e.g., Huster et al., 2014, 2017).

The P3 is consistently increased under conditions conceptually linked to high inhibitory load. Lowered stop-signal probabilities, cue-induced response preparation prior to stop-signal presentation, or faster average response times are all associated with increased P3 amplitudes (reviewed in Huster et al., 2013). In addition, potential impairments of inhibitory control are often associated with decreased P3 amplitudes, as seen with ADHD for example (e.g., Bekker et al., 2005; Lansbergen et al., 2007). It has also been shown that P3 peak latencies in unsuccessful stop trials are delayed relative to successful stop trials (e.g., Kok et al., 2004; Ramautar et al., 2004, 2006). Wessel et al. (2015) suggested that the onset of the P3, rather than its peak amplitude or peak latency, may be a more specific marker of response inhibition, as the timing of the P3 onset coincides with the SSRT. And indeed, a positive correlation was found such that participants with longer SSRTs also exhibited later P3 onsets.

However, it has to be noted that no explicit set of criteria has been established yet that would specify the prerequisites of an EEG marker of inhibitory control. Previous work seems to assume that such a marker should exhibit stronger activity with more efficient inhibitory processing or under conditions of higher inhibitory load. Delayed latencies in unsuccessful as compared to successful stop trials also seem to be considered a relevant criterion, as is the correlation with the SSRT. Already this rather small set of criteria is not without some conceptual issues that need further clarification though. First, experimental manipulations of inhibitory load may often be confounded by other cognitive factors. Decreasing the probability of stop trials makes it more difficult to successfully inhibit a response, but the relative increase in “novelty” of these stimuli may also augment attentional capture. Second, the SSRT is merely an indirect estimate of inhibitory speed, computed as a difference measure from go reaction times and SSDs. Isolated assessments of associations between the SSRT and selected neural measures may therefore be misleading, since the relevance of moderating or mediating third variables remains unspecified. Recent work of Matzke et al. (2019) exemplifies the susceptibility of established SSRT estimates to “contaminants” of behavioral stop-signal data (e.g., the failure to trigger the go or the stopping process; Matzke et al., 2019). It thus still needs to be tested whether the observed correlations of the P3 with SSRT estimates are specific such that they remain of significant size when accounting for correlations with other behavioral parameters (e.g., the go reaction time). A marker of inhibitory control should furthermore exhibit activity that precedes inhibitory effects at the level of cortical representations or effector muscles that finalize response generation. Previous work relied on the SSRT to delineate such a temporal threshold (about 220 msec), but muscle activity recorded from response effectors suggests that the onset of an inhibitory influence can already be seen at around 150 msec post stop-stimulus presentation. Two lines of research indicate this early effects of inhibition: i) partial response electromyography (prEMG) in successful stop trials shows a rapid and efficient deactivation starting around 150 msec (Raud & Huster, 2017); ii) motor-evoked potentials induced by transcranial magnetic stimulation also exhibit decreased corticospinal excitability beginning at about 150 msec post stop.
stimulus presentation (van den Wildenberg et al., 2010). These EMG-based measures thus precede P3 onset latencies as well as usual SSRT estimates by about 70 msec, and thus prepone the time window in which to look for markers of inhibitory control.

This study set out to investigate the applicability and validity of P3-derived measures as markers of response inhibition. We specifically focused on P3 amplitude and latency measures and their relationship with behavioral indices of SST performance. However, to test the specificity of these associations, we also assessed associations of the N2 with behavioral performance measures, as well as the relationship between the N2 and the P3. Here, the N2 refers to a fronto-centrally maximal negative deflection occurring about 200 msec post stimulus presentation. Just as the P3, the N2 is usually larger in stop than in go trials (e.g., Huster et al., 2013). The N2 is believed to be generated in the midcingulate cortex, and to indicate the occurrence of conflicts in information processing (e.g., Dimoska et al., 2006; Donkers et al., 2004; Huster et al., 2010, 2011, 2014; Ramautar et al., 2004). The first section reports a meta-analysis on studies that specifically tested and reported associations between P3-indices and the SSRT. We then reviewed every article, selecting those that reported to have tested hypotheses derived from this meta-analysis on a data set including EEG, electromyography (EMG), and behavioral performance measures of a SST. We focused on the most common quantification methods, i.e., the extraction of peak amplitudes, peak latencies, as well as onset latencies from standard ERPs as well as decomposed EEG, namely component ERPs derived by subject-specific and group-level independent component analysis (SS-ICA and G-ICA, respectively).

2. A meta-analysis of P3/N2–SSRT associations

A systematic literature review was conducted to identify published studies that assessed and reported associations of P3-derived measures with the SSRT. We further assessed N2–SSRT associations to check for the functional specificity of potential P3 relationships. The last date for literature search and data inclusion was the 4th of July, 2018. Table 1 lists relevant articles alongside their ERP parameters and correlations with the SSRT. Dependent measures corresponding to P3 peak or mean amplitude were subsumed under P3 amp; P3 peak lat refers to the quantification of the latency at which the P3 showed its maximum amplitude; P3 onset lat incorporates variables that quantify the onset of the P3. The same grouping was applied to N2-derived measurements.

2.1. Procedure

To generate a starting list of articles that may contain P3–SSRT correlations, several searches were conducted in PUBMED with varying keywords, of which the combination of “stop signal task” and “EEG” produced the largest and most encompassing list. We then reviewed every article, selecting those that reported to have assessed associations between P3-derived variables and SSRTs, and adding further articles to the list based on cross-referencing.

| Article | P3 amp | P3 peak lat | P3 onset lat | N2 amp | N2 peak lat | N2 onset lat | SSRT | Stim | Resp | Comments |
|---------|-------|-------------|--------------|-------|-------------|--------------|-------|------|------|----------|
| Liottie et al. (2005) | n.s. | int | V | M | 10/10 ADHD/HC |
| Johnstone et al. (2007) | n.s. | n.s. | n.s. | n.s. | mean | VA | M | 24 |
| Van Gaal et al. (2009) | .22 | .53 | mean | V | M | 15 |
| Anguera et al. (2012) | .30 | .60 | .30 | .10 | n.a. | V | M | 17 |
| Hughes et al. (2012) | .01/ .47 | .14/ .90 | n.a. | n.a. | mean | V | VA | M | 13 |
| Senderecka et al. (2012) | .51 | .29 | int | V | M | 16 |
| Jones, et al. (2013) | .61/.43 | int | VA | M | 16 SSD 150/SSD 150-250 |
| Huster et al. (2014) | .61 | mean | V | M | 13 |
| Logemann et al. (2014) | /C0 | /C0 | int | V | VA | M | VA | M | 12/10/14/26 |
| Senderecka et al. (2016) | /C0 | /C0 | .31 | .40 | int | V | M | 33 |
| Wessel et al. (2015) | .3 | n.a. | V | M | 10/13 Older adults |
| Dutra et al. (2018) | .52 | int | V | M | 17 |
| Hoptman et al. (2018) | /C0 | /C0 | .41 | med | V | M | 42 |

Table 1 – Overview of studies that report correlations between N2–P3–measures and the SSRT. Abbreviations: amp = amplitude, lat = latency, n.a. = not significant, value not stated, ADHD = attention deficit/hyperactivity disorder, HC = healthy controls, SSRT after mean method, int = SSRT after integration method, VA = visual and auditory stop signal, Resp = response modality, O = oral response, M = manual response.
in already reviewed articles. This way, we identified a total of 16 articles that reported tests of P3-SSRT correlations. If studies reported statistically non-significant correlations without specifying the exact correlation coefficient, the authors were contacted, and, upon provision, the correlation coefficient was included in the analysis; if the non-significant effect could not be specified any further, n.s. was entered. For each of the variables, except for the N2 onset latencies, a meta-analysis of the correlations was conducted using the MedCalc software (www.medcalc.org, version 18.9.1). Most variables exhibited a significant degree of heterogeneity, and we therefore calculated the summary correlation coefficient under the random effects model (DerSimonian & Laird, 1986).

2.2. Results

Although the table is relatively sparsely beset, the data sufficed for meta-analytic assessment for all ERP parameters but the N2 onset latency. Fig. 1 depicts the weighted study coefficients, the summary correlation coefficients, as well as the corresponding confidence intervals (CIs). With respect to ERP amplitudes, neither the P3 nor the N2 exhibited significant

Fig. 1 – Results of the meta-analysis of correlations between N2-/P3-measures and the SSRT. Depicted are the correlation coefficients per study (x-axis), their relative weighting, the summary coefficient as well as the corresponding 95% confidence intervals.
summary correlations with the SSRT (P3 amp: $r = -0.21$, 95%-CI = $-0.41$ to $-0.04$, $I^2 = 74.01$, $p = .22$; N2 amp: $r = -0.01$, 95%-CI = $-0.39$ to $0.37$, $I^2 = 78.58$, $p = .97$). In contrast, the three remaining latency measures all revealed significant summary correlations with the SSRT (P3 peak lat: $r = 0.3$, 95%-CI = $0.09$ to $0.49$, $I^2 = 42.80$, $p < .008$; P3 onset lat: $r = -0.41$, 95%-CI = $-0.21$ to $-0.69$, $I^2 = 81.05$, $p < .04$; N2 peak lat: $r = 0.46$, 95%-CI = $0.18$ to $0.66$, $I^2 = 46.38$, $p < .003$).

2.3. Interim discussion

Current evidence therefore suggests that neither N2 nor P3 amplitudes are consistently correlated with the SSRT. This is somewhat surprising, since a previous review of the EEG literature seems to suggest that conditions of increased inhibitory load coincide with larger P3s (Huster et al., 2013), a finding that usually is mirrored when comparing healthy controls and patient groups with potentially impaired inhibitory control, with the latter showing attenuated P3 amplitudes.

Latency measures, on the other hand, quite consistently show an association with the SSRT with overall medium effect sizes. Both the P3 peak latency as well as the P3 onset latency are in sum positively correlated with SSRTs, such that earlier P3s coincide with shorter SSRTs. This effect would generally be in accordance with the notion that the latency of the P3 may serve as indicator of inhibitory control capabilities (e.g., Wessel & Aron, 2015). However, this effect is not specific for the P3, since the N2 peak latency shows the same association. For now though, it is unclear whether this is the case for the N2 onset as well (only a single study assessed this association and provided a null finding).

Altogether, this opens the possibility that neither N2 nor P3 latencies serve as specific indicators of the temporal dynamics of inhibition proper, but that these associations may rather be driven by the timing of earlier processing stages (e.g., sensory processing) or by general capacity limitations for higher order cognitive processing.

Different kinds of potential moderator variables could potentially influence the associations of interest. SSTs have been set up using different combinations of stimulation and response modalities, for example. In the studies sampled here, the most common combination was the utilization of purely visual stimuli with manual responses. Only three studies used auditory stop stimuli, and only a single study reported the use of verbal responses. However, it has been reported that such factors can influence RTs and SSRTs (van der Schoot et al., 2005). This begs the question of whether pooling correlations across different samples and/or populations is appropriate. However, our data base is insufficient to assess to what degree differences in the experimental design would moderate SSRT–EEG correlations. Other potential moderator variables refer to different procedures in the analysis of the behavioral or EEG data. The SSRT can be estimated in different ways, with the mean or integration method being the most common ones; yet, more recent procedures based on Bayesian modeling suggest that these measures may actually overestimate the stopping latency (e.g., Matzke et al., 2019; Skippen et al., 2019). Similarly, amplitude measures of ERPs can be derived in different ways, e.g., by extracting the peak amplitude or by computing the mean amplitude over larger time frames. The same holds true for latency measures; the P3 onset latency, for example, has been quantified based on the earliest significant difference between go and stop trials in some studies, whereas others may choose to compute it as the time point by which a certain percentage of the peak amplitude is reached. Also, whereas the majority of studies focusing on the electrophysiology of stopping relies on parameters directly derived from scalp EEG, more recent studies often apply data decomposition techniques to better isolate the latent processes underlying specific EEG components, e.g., via principal or independent component analysis (PCA and ICA, respectively).

Future studies should more regularly report associations of various EEG-derived and behavioral performance measures, so that more sophisticated meta-analyses can be conducted that include the assessment of potential moderator variables. Such brain–behavior associations are worth to be reported, even though it might be in a more exploratory manner, as to build a good foundation for summary assessments such as the meta-analysis conducted here. With this in mind, we now proceed to the empirical assessment of ERP–behavior associations.

3. A comparative analysis of brain–behavior correlations of stopping

We now set out to further assess the association of EEG-derived variables and behavioral performance measures as observed in the SST. We will do so predominantly using exploratory analyses. Nonetheless, based on our previous meta-analytic results we can formulate the following expectations: 1) P3 and N2 latency (but not amplitude) measures correlate with the SSRT; 2) since both ERPs show these latency–SSRT associations, suggesting that earlier processing stages may drive these effects, we expect N2 and P3 latencies to be correlated.

To guide our exploratory interpretation of the correlation coefficients we will focus on correlations of $|0.2|$ and above for two reasons: 1) it follows our expectations of the medium effect sizes and their corresponding variation across studies found in the meta-analysis; 2) it compensates for a potential publication bias that is known to cause an overestimation of actual effect sizes. We will now also assess the overall structure of associations between ERP and behavioral variables. This is necessary, since the SSRT is a difference measure (derived from the goRT and SSD), and up to now there is no data that would clarify how the N2/P3 relate to goRT or stopping accuracy.

We extracted peak amplitudes, peak latencies, and onset latencies from EEG. P3 onset latencies were computed in two different ways: 1) the half-amplitude onset latency (1/2 amp. latency), i.e., the earliest time point at which an ERPs amplitude exceeds half of its peak amplitude when moving backwards in time starting at the peak; 2) the differential onset latency (diff. onset latency) according to Wessel and Aron (2015), i.e., the earliest time point at which go and stop trial activity differs significantly from each other. Onset latencies were estimated for the P3 only because of the high single trial variability in case of the N2.
We furthermore compared dependent variables extracted via different data decomposition techniques based on ICA: subject-specific (SS-ICA) and group independent component analysis (G-ICA). Whereas there is no direct indication to believe that standard ERPs and ICA-based ERPs would differ dramatically in this specific context (the SST), this notion has not directly been tested yet. These comparative analyses will indicate whether findings generalize regardless of differences in EEG processing. Please refer to Fig. 2 for a depiction of EEG and component time-courses, and to Table 2 for an overview of dependent variables.

In accordance with Cortex’s transparency guidelines, we report how we determined the sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. The data, analysis scripts, and experimental files are available at https://osf.io/mhy9b/.

3.1. Participants

Thirty-seven right-handed participants between the age of 19 and 35 years took part in the study. None reported a history of psychiatric or neurological disorders; all had normal or corrected-to-normal vision. Data from four participants was discarded due to low performance on the SST. These participants exhibited signs of an extreme “waiting strategy” with very slow response times and low go accuracies (all < 80%), despite blocked and trial-wise feedback to abstain from such response slowing. The final sample consisted of 33 participants (18 female, 15 male; mean age = 26.6 years). All participants gave written informed consent prior to study participation. The study protocol was approved by the institutional review board of the Department of Psychology at the University of Oslo, and followed ethical standards according to the Declaration of Helsinki. The study procedures and analyses were not pre-registered prior to the research being conducted. The analyses were run on a previously conducted study, thus sample size was fixed prior to the conceptualization of this analysis. All data inclusion/exclusion criteria are described in the manuscript and follow general guidelines to ensure sufficient data quality for the analysis of EEG and behavioral data.

3.2. Task

All participants performed both a go/no-go task (GNGT) and a SST in a single session, of which only the SST is of relevance here. Task order was counterbalanced across subjects. The SST lasted for about 30 min. Task presentation was controlled via E-prime 2.0.

Go-stimuli were green arrows that pointed either to the left or to the right (arrowheads of size 3 cm × 3.5 cm). Stop-stimuli were blue arrows of the same size and orientation. The participants were seated at a viewing distance of approximately 80 cm from the screen. All stimuli were presented at the center of a 14.4” monitor at a refresh rate of 60 Hz. Participants were instructed to respond as quickly and accurately as possible to go-stimuli by pressing the space bar on a keyboard, and to press nothing when stop-stimuli appeared.

In accordance with Cortex’s transparency guidelines, we report how we determined the sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. The data, analysis scripts, and experimental files are available at https://osf.io/mhy9b/.

![Fig. 2](image.png)  
**Fig. 2** – ERPs derived from normal EEG analysis, after subject-specific ICA (SS-ICA), as well as group-ICA (G-ICA). Plots exhibit the P3 scalp distribution in the lower right, with an additional topography for the N2 in the lower left in case of EEG; positive and negative values are red and blue, respectively. Topographies are symmetrically scaled to the absolute maximum value with arbitrary units for SS-ICA and G-ICA, and µV in case of EEG. Shaded areas depict the 95% confidence interval around the condition-specific mean activity.

### Table 2 – Descriptive statistics of the dependent measures (means and standard deviations).

| Measure                      | EEG  | SS-ICA  | G-ICA  |
|------------------------------|------|---------|--------|
| **Behavior**                 |      |         |        |
| goRT            | 605 ± 89.01 |       |       |
| usRT            | 531 ± 85.67 | .98    |       |
| SSTK            | 189 ± 35.48 | -.38   | -.40   |
| ACC             | 52.26 ± 2.41 | .88    | .88    |
| prEMG           | 135 ± 17.28 | -.55   | -.49   | .35    |
| **Latency**      |      |         |        |
| goRT            | 605 ± 89.01 |       |       |
| usRT            | 531 ± 85.67 | .98    |       |
| SSTK            | 189 ± 35.48 | -.38   | -.40   |
| ACC             | 52.26 ± 2.41 | .88    | .88    |
| prEMG           | 135 ± 17.28 | -.55   | -.49   | .35    |
| **Amplitude**    |      |         |        |
| SSRT            | 189 ± 35.48 | -.38   | -.40   |
| ACC             | 52.26 ± 2.41 | .88    | .88    |
| prEMG           | 135 ± 17.28 | -.55   | -.49   | .35    |
of the screen. If none of the target stimuli was displayed, a fixation cross was presented instead at the same location.

Participants were instructed to respond as quickly as possible via button press with the thumb of the hand corresponding to the direction of the go-stimulus. In stop trials, the stop-stimulus appeared after the go-stimulus with a short delay (the SSD), instructing the participant to suppress their already initiated response. Stimuli were presented for 100 msec and the SSD adapted according to a tracking procedure aiming at a stopping accuracy of 50%. After successful stop trials, the SSD was increased by 50 msec, whereas it was decreased by the same time after unsuccessful stop trials. The minimum and maximum SSD were set to 100 and 800 msec, respectively. The SSD tracking was done separately for the left and right hand. The inter-trial interval was randomly varied between 1500 and 2500 msec.

The task consisted of 800 trials, of which 600 were go trials and 200 stop trials, with an equal number of left- and right-hand trials. Blocks of 80 trials were followed by a feedback procedure instructing participants to respond faster if the average go reaction time of the preceding block exceeded 500 msec. Instantaneous feedback (“Too slow!”) was given after a go omission or if the reaction time exceeded 1000 msec. Prior to the SST, participants completed a short training session of 20 stop trials. The latency of the peak of prEMG activity thus informs us about the time at which action cancellation processes effectively result in a reduction of motor activity after action generation (for more info please also refer to Raud & Huster, 2017). To compute the prEMG latency, EEG channels were filtered between 10 and 200 Hz, resampled to 500 Hz, and segmented relative to the stop-stimulus. Trials with amplifier saturation were discarded from the analysis. After computing the root mean square for each time point, a moving average with a window width of 11 data points was applied. The time-series of each trial was then transformed through division by the trial-specific average of pre-go activity from −200 to 0 msec. The single-trial data were then z-scored across all trials and time-points, separately for each hand. The prEMG peak latency was calculated from the averaged time-series of all successful stop trials.

EEG channels were filtered between .1 and 80 Hz, resampled to 500 Hz, and re-referenced to the common average reference computed over all EEG channels. Infomax independent component analysis (ICA) was run on the data using the routines provided in EEGLAB, and components capturing eye or muscle artifacts were identified and rejected manually. Data were then subjected to another low-pass filter at 40 Hz, and epochs from −200 to 800 msec relative to the go- and stop-stimulus were extracted for valid go trials, as well as successful and unsuccessful stop trials, respectively. A baseline-correction was computed using the −200 to 0 msec interval and subtracting the baseline average from the whole time series of a trial. An automatic artefact rejection algorithm was run as implemented in EEGLAB’s pop_autorej-function, and the remaining epochs were visually inspected for residual artifacts to correct those manually.

To identify the P3-component based on SS-ICA, an ICA was run on the cleaned data with the number of extracted components equal to the number of EEG electrodes minus the number of components rejected during artifact correction for a given data set. The resulting components were then inspected and the component capturing the P3 (based on topography and time course) was selected for further processing.

To compute a G-ICA, 100 go trials as well as 60 successful and 40 unsuccessful stop trials were randomly selected from each data set (with equal contribution of left- and right-hand trials). These numbers were based on the minimum amount of trials available across data sets, while still allowing for the calculation of reliable ERPs, since the organization of G-ICA relies on the concatenation of equally-sized and structured data sets. G-ICA concurrently estimates a component structure representative for the whole group of data sets by combining subject-specific and group-level PCA with a group-level ICA (for details, please refer to Eichele et al., 2011; or Huster et al., 2015, 2018). Here, we extracted a total of 8 components, since the first-level PCA indicated that 8 components explained about 90% variance in each of the single-subject data sets. We then identified the group-component that captured the P3 (based on the component time course and topography). This component was reconstructed for each single data set by extracting the subject-specific demixing matrix and applying it to the single-trial EEG data of each

3.3. Data acquisition

EEG and EMG were recorded using a Neuroscan SynAmps2 amplifier with a sampling rate of 2500 Hz, an online high-pass filter at .15 Hz, and an online low-pass filter at 1000 Hz. EEG was measured from 64 passive Ag/AgCl electrodes placed in accordance with the extended 10–20 system with two additional horizontal electrooculography (EOG) channels placed beside the left and the right eye. All EEG electrodes were referenced online against a nose-tip electrode. Impedances were kept under 5 kOhm. For the EMG, the same type of Ag/AgCl electrodes were used in bipolar recording schemes with placements above the abductor pollicis brevis. The ground electrode was placed on the left arm. The participants’ arms were supported using pillows to reduce spurious baseline muscle tension.

3.4. Analyses

Go trial reaction times (goRT), stopping accuracies, unsuccessful stop trial reaction times (usRT), errors of commission and omission in go trials, as well as the SSRT were computed. The SSRTs were estimated separately for left- and right-hand responses using the integration method, i.e., by subtracting the mean SSD from the go-reaction time distribution percentile corresponding to the probability of unsuccessful stopping. Trials with go omissions or erroneous responses were excluded from the computation of the SSRT. All behavioral measures will be reported after averaging across both hands.

Another “behavioral” index of stopping, the partial response EMG (prEMG) activity in successful stop trials, was derived from the EMG recordings. PrEMG reflects the muscle activity of the task-relevant effector muscles in successful stop-trials. The latency of the peak of prEMG activity thus informs us about the time at which action cancellation processes effectively result in a reduction of motor activity after action generation (for more info please also refer to Raud & Huster, 2017). To compute the prEMG latency, EEG channels were filtered between 10 and 200 Hz, resampled to 500 Hz, and segmented relative to the stop-stimulus. Trials with amplifier saturation were discarded from the analysis. After computing the root mean square for each time point, a moving average with a window width of 11 data points was applied. The time-series of each trial was then transformed through division by the trial-specific average of pre-go activity from −200 to 0 msec. The single-trial data were then z-scored across all trials and time-points, separately for each hand. The prEMG peak latency was calculated from the averaged time-series of all successful stop trials.

The SSRTs were estimated separately for left- and right-hand trials. Blocks of 80 trials were followed by a feedback procedure aiming at a stopping accuracy of 50%. After successful stop trials, the SSD was increased by 50 msec, whereas it was decreased by the same time after unsuccessful stop trials. The minimum and maximum SSD were set to 100 and 800 msec, respectively. The SSD tracking was done separately for the left and right hand. The inter-trial interval was randomly varied between 1500 and 2500 msec.
subject, thereby reconstructing all available go and stop trials at the component level (a more detailed description as well as example code can be found in Huster et al., 2018).

ERPs were computed for correct go trials, as well as for successful and unsuccessful stop trials for i) the normal EEG data, ii) the subject-specific P3 components extracted using SS-ICA, and iii) the P3 components reconstructed for each subject using G-ICA. Thus, these ERPs differed in their pre-processing, but otherwise contained the exact same trials. We extracted the N2 peak amplitude and peak latency, as well as the P3 peak amplitude, peak latency, \( \frac{1}{2} \)-amplitude onset latency, as well as the differential onset latency. The basic time windows were defined as 150—300 msec, and 225—420 msec for the N2 and P3 measures, respectively. The time window for the P3 was adapted so that the P3 diff. onset latency could be estimated reliably (see below for details). The N2/P3 were defined as the most negative/positive value within respective time windows. The peak amplitude was calculated as the mean amplitude at peak \( \pm \frac{1}{2} \) 5 data points, and the peak latency simply as the latency of the local minimum/maximum relative to stop-stimulus onset. The P3 \( \frac{1}{2} \)-amplitude onset was computed as the time point at which the amplitude first reached a value 50% smaller than peak amplitude when tracing amplitudes backward in time starting at the peak. To compute the P3 diff. onset latency, go and stop trials of the same SSD were matched to control for the influence of motor preparation, and then permutation based statistics were used to determine the earliest time point at which go and stop trial activity differed significantly from each other (for details please refer to Wessel & Aron, 2015). An average lower boundary of 225 msec was established for the search window by visually inspecting the single-subject ERPs as well as their corresponding onset estimates. The P3 diff. onset latency could not be estimated in one subject based on the regular EEG data; this subject’s P3 diff. onset latency was thus set to the mean value of the sample. Single-trial N2 and P3 amplitudes were computed by extracting for each trial the mean amplitude of a 100 msec time window centered on each individual’s N2 or P3 peak latency.

The signal-to-noise-ratio (SNR) was computed for each of the three approaches by dividing the root-mean-squared amplitude values of a 20 msec time-window centered on the P3 peak by the root-mean-squared amplitude values of the 200 msec baseline period. The SNR was computed based on successful stop trials only.

Correlation coefficients were computed as standard bivariate product moment correlations. The correlations between the dependent variables were then visualized by means of graph construction. First, simple graphs were computed for each of the data processing methods by setting a specific threshold for the correlations. Correlations exceeding this threshold (as \( |r| \)) contributed edges to the graph, whereas the behavioral or EEG variables constituted the nodes. We then integrated the structures of the graphs derived for each of the three methods by computing a multigraph that thus could contain up to three edges between each pair of nodes. At last, a simple graph was computed by keeping only those edges between two nodes of the multigraph that were found in at least two of the three analysis approaches. This procedure was repeated with four different thresholds \( (r = .2, .3, .4, \) and .5) to highlight the graph structure and its change based on effect size.

3.5. Results

3.5.1. Behavioral data

Overall, behavioral performance measures were within the normal range of what would be expected of a plain visual SST with an average goRT of 605 msec, an SSRT of 189 msec, and a stopping accuracy of 52%. The percentage of correct go responses was high at 94%. The percentage of erroneous responses to go stimuli (e.g., pressing the right button when left would have been correct) was lower than .6%, and the number of go omissions was acceptable at 5.4%. The mean SSD was 400 msec. RTs for unsuccessful stop trials (usRT: 531 msec) were significantly shorter than normal goRTs \( (t_{32} = -22.05, p < .001) \); this was also the case for every single participant. The behavioral performance measures also exhibited substantial inter-correlations, which are listed in Table 2 together with other descriptive statistics.

3.5.2. EEG and decomposed ERPs

Fig. 2 depicts the ERP and component time-courses and topographies reflecting the N2 and P3. The depicted N2/P3-complex is most pronounced at fronto-central areas of the scalp. Relative to the EEG-ERPs, both ICA procedures seem to dissociate the N2/P3-complex from other EEG phenomena. This can, for example, be seen with the N2, which shows some spatio-temporal overlap with activity over occipital areas, or with the slight drift in the baseline EEG of stop trials. This dissociation of different sources is also reflected in nominally higher mean SNRs for the ICA procedures as compared to standard EEG processing (EEG: 15.13; SS-ICA: 19.83; G-ICA: 20.21).

P3 peak amplitudes were larger for unsuccessful stop trials with all three analysis methods (EEG: \( t_{32} = -3.34, p < .01 \); SS-ICA: \( t_{32} = -3.89, p < .001 \); G-ICA: \( t_{32} = -4.34, p < .001 \)), and P3 peak latencies were significantly later (EEG: \( t_{32} = -6.98, p < .001 \); SS-ICA: \( t_{32} = -4.35, p < .001 \); G-ICA: \( t_{32} = -3.78, p < .001 \)). N2 amplitudes were significantly larger (more negative) with EEG \( (t_{32} = 3.41, p < .01) \), but not with SS-ICA \( (t_{32} = 1.14, p = .26) \) or G-ICA \( (t_{32} = 1.58, p = .12) \). N2 peak latencies were also delayed in unsuccessful relative to successful stop trials (EEG: \( t_{32} = -3.2, p < .01 \); SS-ICA: \( t_{32} = -4.05, p < .001 \); G-ICA: \( t_{32} = -3.25, p < .01 \)).

3.5.3. P3/N2—SSRT associations

The data indicated that both N2 and P3 latencies are associated with the SSRT. Please refer to Table 3 for the exact correlation coefficients. P3 peak and onset latencies obtained from EEG and ICA-decompositions overall showed medium-sized correlations with the SSRT in the expected direction, such that later P3 peaks were associated with longer SSRTs. The same pattern emerged for the N2 peak latency (except for the G-ICA-based latency measure), with shorter latencies corresponding to shorter SSRTs.

P3 amplitudes did not correlate highly with the SSRT; EEG-derived N2 amplitudes exhibited a small-to-medium negative correlation though: larger (i.e., more negative) ERPs were associated with longer SSRTs.
3.5.4. Exploratory P3/N2-behavior correlations

We conducted further exploratory correlational analyses between the ERP-derived amplitude and latency measures on the one hand, and goRT, stopping accuracy, and prEMG activity in successful stop trials on the other hand. Most studies assess or report correlations hypothesis-driven and thus selectively, usually focusing on the SSRT. However, since the SSRT is a difference measure derived from goRTs and SSDs, it is important to also inspect the overall correlation structure. Table 3 lists the correlation coefficients.

These exploratory analyses indicated that P3 amplitudes and P3 onset latencies are related to both the average goRT and stopping accuracy. Larger P3 amplitudes and earlier onset latencies co-occurred with longer goRTs and higher stopping accuracy. These effects were, however, less pronounced with the ICA-based \( \frac{1}{2} \) amplitude latency measures. Similar effects were found for the N2. The N2 peak latency was associated with goRTs and stopping accuracies, at least when extracted via EEG or SS-ICA, and N2 amplitudes derived via G-ICA correlated negatively with goRTs. As with the P3, earlier N2 latencies and larger N2 amplitudes (i.e., more negative) were associated with longer goRTs and higher stopping accuracies. With respect to the peak latency of the prEMG activity in successful stop trials, N2 and P3 latency measures largely exhibited positive correlations. Thus, later peak latencies of prEMG activity in successful stop trials were associated with later N2 and P3 latencies.

3.5.5. Graph estimation and visualization

Fig. 3 depicts the simple graphs that assess the structure of the dependencies between the behavioral and EEG-derived variables across the different data processing methods for the four different thresholds \( r = \{0.2, 0.3, 0.4, 0.5\} \). Although it could be expected that the variables show clusters according to their modality (behavior vs EEG), it is interesting to note that the EEG-derived variables further break up into two clusters with medium to high correlations coefficients, namely those quantifying amplitudes and those specifying latencies. It further seems that these clusters show differential associations with behavioral markers. Whereas the amplitude measures (especially P3amp) are predominantly associated with reaction time measures (usRT, goRT), the EEG latency measures (especially N2 and P3 peak latencies) exhibit stronger associations with the prEMG latency, the stopping accuracies, as well as the goRT. Overall, associations between electrophysiological parameters and the SSRT are weaker than those with the other behavioral variables, suggesting that correlations between EEG-derived variables and the SSRT may be mediated through other behavioral variables.

### Table 3 – Correlations between the ERP-derived latency and amplitude measures and the SSRT, goRT, stopping accuracy, and stopping EMG. ERP amplitudes and latencies were extracted from successful stop trials.

| Variable                                      | EEG    | SS-ICA | G-ICA |
|-----------------------------------------------|--------|--------|-------|
| **SSRT correlations**                        |        |        |       |
| P3 amplitude                                  | -0.19  | 0.02   | -0.02 |
| P3 peak latency                               | 0.19   | 0.30   | 0.24  |
| P3 \( \frac{1}{2} \) amp. onset latency       | 0.29   | 0.25   | 0.25  |
| P3 diff. onset latency                        | 0.21   | 0.12   | 0.25  |
| N2 amplitude                                  | -0.28  | -0.11  | -0.08 |
| N2 peak latency                               | 0.26   | 0.21   | 0.02  |
| **goRT**                                      |        |        |       |
| P3 amplitude                                  | 0.48   | 0.25   | 0.48  |
| P3 peak latency                               | -0.14  | -0.20  | -0.21 |
| P3 \( \frac{1}{2} \) amp. onset latency       | -0.35  | -0.22  | -0.08 |
| P3 diff. onset latency                        | -0.50  | -0.32  | -0.37 |
| N2 amplitude                                  | 0.12   | -0.34  | -0.36 |
| N2 peak latency                               | -0.42  | -0.51  | -0.11 |
| **Stopping accuracy**                         |        |        |       |
| P3 amplitude                                  | 0.37   | 0.13   | 0.38  |
| P3 peak latency                               | -0.17  | -0.23  | -0.20 |
| P3 \( \frac{1}{2} \) amp. onset latency       | -0.31  | -0.20  | -0.09 |
| P3 diff. onset latency                        | -0.50  | -0.32  | -0.46 |
| N2 amplitude                                  | 0.13   | -0.32  | -0.28 |
| N2 peak latency                               | -0.32  | -0.47  | -0.09 |
| **Partial response EMG**                      |        |        |       |
| P3 amplitude                                  | -0.17  | -0.03  | -0.24 |
| P3 peak latency                               | 0.42   | 0.53   | 0.40  |
| P3 \( \frac{1}{2} \) amp. onset latency       | 0.46   | 0.36   | 0.29  |
| P3 diff. onset latency                        | 0.25   | 0.31   | 0.36  |
| N2 amplitude                                  | -0.04  | 0.26   | 0.24  |
| N2 peak latency                               | 0.64   | 0.50   | 0.08  |

N2 and P3 peak latencies exhibited relevant positive correlations with each other with \( r = 0.39 \) for EEG, \( r = 0.55 \) for SS-ICA, and \( r = 0.26 \) for G-ICA. Similarly, N2 and P3 amplitudes were negatively correlated when extracted via SS-ICA (\( r = -0.49 \)) and G-ICA (\( r = -0.63 \)), such that more negative going N2s co-occurred with larger P3s. For EEG-derived amplitude measures, the correlation was found to be \( r = 0.02 \) only.

4. Discussion

Both the meta-analytic as well as the empirical data replicate the previously reported associations between the P3 latency and the SSRT with small to medium effect sizes. However, this association is not specific, neither with respect to its underlying EEG phenomenon, the P3 as compared to other ERPs, nor regarding its association with the SSRT. Considering the overall pattern of results, we conclude that neither the P3 nor the N2 fulfill a minimal set of criteria necessary to qualify as potential marker of inhibitory control.

4.1. Replicability of results using different EEG-analysis schemes

First, it is important to point out that the patterns of associations between neural and behavioral markers were very similar regardless of whether ERP measures were derived directly from the EEG, or from data decomposed via subject-specific or group-level ICA. For the P3-derived measures, relevant associations were of similar size and direction across...
methods. This does not mean though that the correspondence is perfect and that these methods can be used interchangeably. Higher SNRs of potentials were obtained from ICA procedures, which is of importance for analyses based on single-trial data. On the other hand, comparisons of component strengths for a specific condition across subjects using SS-ICA is usually hampered by the independent scaling of component time-courses and weights. G-ICA minimizes this problem, but suffers from decreased sensitivity to source patterns with rather weak time-locking (see Huster et al., 2015, for an in-depth discussion). Thus, when interested in group-comparisons or analyses across subjects, G-ICA seems better suited than SS-ICA, which again might be a better choice for the optimal reconstruction of an individual’s component structure including activity patterns with poor time-locking; standard EEG analyses seem to be the compromise, as they are limited with respect to signal quality, especially when single-trial data are of interest.

4.2. EEG markers and their association with the SSRT

An EEG marker of inhibitory control can be expected to exhibit correlations with the current best-established marker of the stopping latency, namely the SSRT. These associations further should be of relevant size, and optimally they would be specific. Specificity, i.e., that no other EEG phenomena show the same association, may be considered a conservative criterion, yet in our perspective it is also a necessary one. The SSRT is not a pure measure of the performance of the stopping system in so far as it includes several processing stages. Let’s just consider a sensory processing stage, as well as the actual cycle through the inhibitory system, both of which exhibit individual differences. An association of the SSRT with P3 latency measures can thus (at minimum) be driven by individual differences in inhibitory capabilities, sensory processing, or both. The P3, because of its temporal co-occurrence with the presumed stopping latency and based on its association with the SSRT, has been suggested to be a marker of response inhibition, yet not sensory processing. This functionally specific claim necessitates that we rule out that our P3–SSRT correlations are driven by factors other than inhibition proper. The fact that both the meta-analysis as well as our empirical data also indicate associations between the SSRT and the N2 latency, which precedes the P3 by about 150 msec, undermines such a functionally specific interpretation. Thus, future studies need to specifically address whether SSRT–P3 associations are inhibition-specific, or whether earlier processes different from inhibition proper drive these associations.

The meta-analysis further indicate that, although both P3 and N2 latency measures exhibit associations with the SSRT, N2 and P3 amplitudes do not seem to be significantly associated with the stopping latency. This pattern suggests a certain...
degree of differential functional specificity of the latency and the amplitude measures. The systematic literature review and the meta-analysis also suggest that only few studies assess ERP–SSRT associations, and of those not all report the actual correlation coefficient if statistical testing did not indicate significance. This, and the fact that publication bias probably is prevalent also in the neuroscience literature, suggests that the actual effect size might be smaller than estimated here.

The empirical analysis of associations of N2 and P3 latency and amplitude measures with behavioral parameters further supports the notion that the P3–SSRT association is less specific than thought. This association is not specific to the P3 onset estimated from matched go and stop trials (e.g., Wessel et al., 2015), but it can similarly be found for the P3 ½-amplitude onset latency as well as the peak latency. Furthermore, the association of the SSRT with the N2 peak latency is of similar size. The fact that many of the associations we find between the P3-derived and behavioral measures also seem to be existent with N2-derived measures further undermines the hope that the P3 may serve as specific markers of inhibitory control.

But even if we were to accept a certain degree of specificity of P3 latency/SSRT associations, other conceptual issues still remain. Correlations of small to medium size leave the much bigger part of the variation unexplained; with correlations of about .3, we achieve less than 10 percent of explained variance. This seems insufficient to consider the here observed latency measures reliable markers of the stopping latency or inhibitory control, or the P3 a “motor inhibition component” (e.g., Dutra et al., 2018).

4.3. EEG markers and their association with other behavioral parameters

Another aspect related to the functional specificity of EEG–SSRT correlations is the pairwise association with other behavioral parameters. Given that the SSRT is derived from SSDs and their specific stopping accuracies as well as goRTs, we have to ensure that correlations with the SSRT are not merely driven by the variance of its constituent variables. Larger correlations of an EEG marker with behavioral parameters other than the SSRT may indicate such a situation, yet this conclusion is admittedly hindered by the lower reliability of the SSRT that necessarily inherits the unreliability of its constituent measures (e.g., the SSDs and goRTs). It is nonetheless crucial to inspect the overall pattern of correlations to assess whether an otherwise hidden process may drive the EEG–SSRT correlations. Yet, our review of published studies indicates a marked tendency for exclusively hypothesis-driven and thus selective testing and/or reporting of effects. Even studies that specifically aimed at the assessment of brain–behavior associations usually did so by merely testing the effect of interest (e.g., the association of P3 latency and SSRT). Whereas this procedure is commendable in its goal to minimize false positive findings, it comes at the risk of missing potential mediator or moderator variables.

Indeed, our own empirical data show that correlations of EEG markers with the SSRT again are not specific, but associations of ERP latency measures with the goRT, stopping accuracy and prEMG are at least of similar size. The graph visualization highlights the somewhat isolated positioning of the SSRT at higher effect sizes, which may suggest that its associations with EEG-derived measures is mediated via other behavioral parameters or an underlying latent process. This interpretation is also strengthened by the finding that P3 latency–SSRT correlations are further attenuated when controlling for goRTs (from an average correlation of .23 to .14). Whereas this procedure may also correct for common factors such as general or sensory processing speed, one should still expect relevant associations of a potential marker of inhibition with the SSRT after this correction given that the stopping latency is considered independent of go-processes. Current work suggests that the prEMG may serve as a more direct marker of the stopping latency, as it does not rely on calculations related to the goRT (e.g., Raud & Huster, 2017). Yet, the prEMG peak latency shows similar associations with the goRT (and stopping accuracy, for that matter), which thus suggests that indeed some latent process may cause this pattern of correlations between the SSRT, prEMG, goRT, and stopping accuracy. We will come back to this issue at a later point in the discussion.

4.4. The processing cascade of stopping

One also needs to consider the temporal positioning of a marker of inhibitory control relative to the stopping latency. Necessarily, a marker of inhibitory control must precede an estimate of the stopping latency, or more precisely the onset of inhibitory effects. For this study, the SSRT would place such a temporal threshold at 189 msec, whereas the prEMG estimates the onset of inhibitory effects at 135 msec. Despite of their correlations with both the SSRT and the prEMG, none of the latency measures of the P3 or N2 would thus fulfill this criterion as these all were estimated to occur after 200 msec.

Whereas every selection of latency measures has to be incomplete given the multitude of ways to estimate ERP latencies (e.g., Kiesel et al., 2008), our selection of latency measures reflected the predominance of some measures in the current literature. Nonetheless, no P3 latency estimate would precede the lower boundary established by the prEMG. The N2 peak latency also does not precede this threshold, although one might argue that N2 onset latencies might do so. Whereas we could not compute the N2 onset based on single-trial data due to the lower SNR (as for the P3 diff. onset latency), we calculated the half-amplitude onset latencies for the N2 to be at 159 msec, 176 msec, and 187 msec for the EEG-, SS-ICA-, G-ICA-derived measures, respectively. Overall, the data confirm our conclusion that both the P3 and N2 are unlikely to be valid candidates of inhibitory control, as they do not pre- but succeed the onset of inhibitory effects measured via the prEMG.

4.5. EEG markers of stopping in successful and unsuccessful stop trials

It has been argued that a potential EEG marker of stopping should be delayed in unsuccessful relative to successful stop trials, and that it may also be attenuated in the former case.
And indeed, the latencies for both the P3 and N2 were later for unsuccessful stopping in our data. P3 amplitudes were enlarged for unsuccessful stop trials though. However, whereas smaller P3 amplitudes for unsuccessful compared to successful stop trials have indeed been reported in previous studies (e.g., Greenhouse & Wessel, 2013), the results reported here are not unique. Kok et al. (2004), for example, also found larger P3s in unsuccessful stop trials. Kok et al. further reported that the differential association of successful and unsuccessful stop trials with shorter and longer SSDs, respectively, can cause or at least bias P3 amplitude differences between these trial types. Correspondingly, Hsieh and Lin (2017) visualized ERPs for successful and unsuccessful stop trials (Figures 3 and 4); and although these two conditions were not formally compared, their data suggest that correcting for the temporal overlap of go- and stop-related ERPs at least seems to attenuate amplitude differences.

4.6. Synopsis and hypothesis generation

4.6.1. Criteria for direct and indirect markers of inhibitory control

Previous work has not explicitly defined the criteria against which one should test potential markers of inhibition. While defining such criteria, we further propose to differentiate direct from indirect markers of inhibitory control. Direct EEG markers would correspond to the neural signature of inhibitory control itself. An indirect marker would not correspond to inhibitory processes per se, yet may still be associated with them through other mechanisms. For example, a process monitoring the efficacy of behavioral performance, or inhibitory efficacy for that matter, might still covary in its timing with the monitored process. This distinction between indirect and direct markers is of great importance, because a direct marker i) may enable us to deduct the occurrence of inhibition from its presence in different tasks, and ii) constitutes a target for procedures such as transcranial electrical or magnetic stimulation to increase inhibitory capabilities. A direct marker of inhibitory control may thus aid the design of diagnostic instruments and interventions. We propose the following minimal set of criteria for a direct marker of inhibitory control: 1) a direct EEG marker should reflect the activity of neurons tied to this specific process; 2) such a marker should be delayed and/or attenuated in unsuccessful relative to successful stop trials (assuming the validity of the horse race model); 3) it temporally precedes the effects of inhibition manifest at the motor cortex or the effector muscles; 4) the association with behavioral markers of inhibition is not mediated by other factors. In addition, like every other psychometric tool, an EEG marker of inhibitory control would have to fulfill basic psychometric properties related to objectivity, reliability, and validity.

In sum, we propose that neither the P3 nor the N2 can be considered direct markers of inhibitory control. The single criterion these ERPs seem to fulfill is based on them being delayed in unsuccessful relative to successful stop trials. Our data also indicate problems with psychometric properties. The overall low correlations of the EEG measures with the SSRT indicate low construct validity, which is in accordance with our previous conclusion. However, low correlations may of course also result from low reliability of the latency measures at hand, or from low reliability or validity of the construct criterion (e.g., the SSRT). Future studies proposing markers of inhibitory control also need to more stringently test these psychometric aspects.

4.6.2. To stop or to go: performance monitoring in the SST

As neither the P3 nor the N2 qualify as direct marker of inhibitory control, what underlying factor may then drive the correlations with the SSRT and prEMG? The constellation of correlations observed here is compatible with the notion that these ERP–SSRT associations may relate to other factors such as performance monitoring and behavioral adaptations.

First, we need to consider whether the correlations of ERP latency measures with the SSRT may merely be a byproduct of their associations with other behavioral variables; after all, the SSRT only indirectly estimates the stopping latency based on goRTs, SSDs and stopping accuracies. Differences in the weighting of stopping and going or the speed-accuracy trade-off, for example, may be factors that change the SSRT via its reliance on the goRT and stopping accuracy. Our data indeed indicate that participants exhibiting high stopping accuracy also show longer response times and shorter SSRTs. This is in line with experimental evidence showing that, also within subjects, motivating correct stopping prolongs RTs, increases stopping accuracy, and lowers SSRTs, whereas fast responding leads to an inversion of these effects (e.g., Greenhouse & Wessel, 2013; Leotti & Wager, 2010). However, it cannot fully be excluded that this pattern is driven by a true change in the stopping latency associated with the switch from a reactive to a more proactively controlled stopping mode when correct stopping is stressed. The prEMG as an estimate of the stopping latency is at least in its calculation independent of goRTs, and in our data shows the same pattern of associations across subjects such that shorter prEMG latencies are related to longer goRTs and higher stopping accuracies. Thus, although future studies need to test whether SSRT and prEMG estimates of the stopping latency indeed show similar modulations through the manipulation of motivational factors within subjects, this contradicts the assumption of a merely artificial EEG–SSRT association driven by goRT-related variance.

Is it possible then that the P3/N2–behavior associations are driven by more general performance monitoring operations? N2- and P3-like waves can be observed across a variety of different tasks and contexts: after stop or incongruent stimuli, feedback signals as well as errors. The N2, error-related negativity and the feedback-related negativity, for example, all seem to originate from the midcingulate cortex and have been associated with conflict monitoring, prediction errors, and the processing of novel stimuli (e.g., Gruendler et al., 2011). Similarly, the P3a, P3b, error and feedback positivity can also be considered members of the same family of ERPs. Whereas traditionally the P3a and P3b were often viewed as somewhat functionally dissociated according to the context updating theory (e.g., Polich, 2007), more recent accounts loosen this differentiation and rather stress their common role in information integration. Barcelo et al. (2018) suggested that the P3 reflects the informational content stimuli have relative to the local and global task context. Predictions formalized through information theory were found to hold in
task switching, odd-ball, and go/no-go tasks. Similarly, Murphy et al. (2015) found that the P3 in general mirrored evidence accumulation as estimated via a drift-diffusion model, and further that both fronto-central theta activity as well as the P3 were associated with error awareness in both go and no-go trials. These studies therefore strongly suggest that the P3 is associated with functions other than inhibition. It is furthermore well-documented that these performance monitoring processes relate to behavioral adjustments after errors as well as conflict-laden and feedback stimuli; classical examples being post-error slowing or the conflict adaptation effect (see Ullsperger et al., 2014, for an in-depth review).

In sum, the association of the SSRT with the P3- and N2-derived measures seems to be mediated via processes associated with the weighting of going and stopping. This effect does not merely appear to be an artefact of the subtraction method underlying the SSRT, as we find a similar pattern for the prEMG. Rather, we hypothesize that processes associated with performance monitoring and behavioral adaptation reflected in the N2 and P3 drive this association.

4.6.3. Functional specificity of ERP amplitudes and latencies

Lastly, it seems noteworthy to discuss whether our data may hint at a potential functional dissociation of EEG amplitude and latency measures, and to what degree the effects found here may depend on earlier processing stages. The meta-analysis indicated significant summary statistics for the latency but not the amplitude measures. A visual inspection of the graphs in Fig. 3 may further suggest that amplitude and latency measures exhibit a certain degree of differential clustering. However, it needs to be stressed that the strengths of the summary associations of ERP and behavioral measures do not show major differences between amplitudes and latencies relative to the variation across analysis methods (see Table 3). Data of a recently published study by Skippen et al. (2019) seems to be in accordance with this notion. The authors found relevant correlations of the P3 amplitude and the N1 amplitude with the SSRT. Reported effect sizes were in the range one could expect based on our meta-analysis, with absolute correlations around .3. Thus, Skippen et al. found that relevant correlations with the SSRT occurred already earlier than P3 onset or peak latency, namely with the N1. The N1 peaked at around 132 msec and is generally believed to reflect attentional processing. The data of Skippen et al. thus corroborate our hypothesis that correlations between the P3/N2 and the SSRT do not necessarily reflect a direct correspondence to the cortical act of inhibition, but rather the efficacy of early sensory processing stages.

4.7. Conclusion

The P3 has regularly been implicated in inhibition, not least based on reports that especially its onset latency correlates with the SSRT. Through meta-analysis and empirical data we found this correlation to be replicable, yet also unspecific. Similar correlations can be found for the P3 as well as the N2 peak latencies. Even more, correlations of the same ERP indices were higher for other behavioral indices such as the go reaction time and stopping accuracy. We conclude that it is unlikely that the reported associations between the P3 and the SSRT indeed reflect genuine inhibitory control. Rather, these associations may result from more general behaviorally adaptive patterns such as the weighting of going and stopping or general behavioral adaptations as a consequence of performance monitoring operations.

CRediT author statement

René J. Huster: Conceptualization, Formal Analysis, Writing – Original Draft.
Mari S. Messel: Data Collection, Data Curation, Writing – Review and Editing.
Christina Thunberg: Data Collection, Data Curation, Writing – Review and Editing.
Liisa Raud: Data Collection, Data Curation, Formal Analysis, Writing – Review and Editing.

Open practices

The study in this article earned an Open Data badge for transparent practices.

Acknowledgments

This work was supported by research funds provided by the Department of Psychology at the University of Oslo. We would further thank Magdalena Senderecka, Matthew Hughes, and Alexander Logemann for providing additional data in support for the meta-analysis.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2020.05.021.

References

Anguera, J. A., & Gazzaley, A. (2012). Dissociation of motor and sensory inhibition processes in normal aging. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 123(4), 730–740. https://doi.org/10.1016/j.clinph.2011.08.024
Aron, A. R. (2007). The neural basis of inhibition in cognitive control. The Neuroscientist, 13(3), 214–228. https://doi.org/10.1177/1073858407299288
Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. Trends in Cognitive Sciences, 18(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003
Band, G. P. H., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. Acta Psychologica, 112(2), 105–142. https://doi.org/10.1016/S0001-6918(02)00079-3
Barceló, F., & Cooper, P. (2018). An information theory account of late frontoparietal ERP positivities in cognitive control. Psychophysiology, 55(3),
Murphy, P. R., Robertson, I. H., Harty, S., & O’Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. Elife, Article e11946.

Nigg, J. T., Carr, L., Martel, M., & Henderson, J. M. (2007). Concepts of inhibition and developmental psychopathology. In D. S. Gorfein, & C. M. MacLeod (Eds.), Inhibition in cognition (pp. 259–277). Washington, DC, US: American Psychological Association.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128–2148.

Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: The N2/P3 complex further validated. Brain and Cognition, 56(2), 234–252. https://doi.org/10.1016/j.bandc.2004.07.002

Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2006). Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. Biological Psychology, 72(1), 96–109. https://doi.org/10.1016/j.biopsycho.2005.08.001

Raud, L., & Huster, R. J. (2017). The temporal dynamics of response inhibition and their modulation by cognitive control. Brain Topography, 30(4), 486–501. https://doi.org/10.1007/s10548-017-0566-y

Senderecka, M. (2016). Threatening visual stimuli influence response inhibition and error monitoring: An event-related potential study. Biological Psychology, 113, 24–36. https://doi.org/10.1016/j.biopsycho.2015.11.003

Senderecka, M., Grabowska, A., Szewczyk, J., Gerc, K., & Chmylak, R. (2012). Response inhibition of children with ADHD in the stop-signal task: An event-related potential study. International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 85(1), 93–105. https://doi.org/10.1016/j.ijpsycho.2011.05.007

Skippen, P., Fulham, W., Michie, P., Matzke, D., Heathcote, A., & Karayanidis, F. (2019). Reconsidering electrophysiological markers of response inhibition in light of trigger failures in the stop-signal task. bioRxiv, Article 658336. https://doi.org/10.1101/658336

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. Physiological Reviews, 94(1), 35–79.