Registered Report, Stage 1 (2025), in-principle accepted at *Cortex*

Revisiting the link between error significance and the error-related negativity: A multi-lab direct replication

Magdalena Senderecka [1]*, Jarosław M. Michałowski [2], Łukasz Krzywoszański [3], Christian Bellebaum [4], Christine Albrecht [4], Paweł Basoń [1,5], Piotr Buczkowicz [3,5], Agnieszka Fudali-Czyż [3], Anna Grabowska [1,5], Patrycja Kałamała [1], Filip Sondej [1], Ewa Wiwatowska [2], Greg Hajcak [6], Faisal Mushtaq [7] & Yuri G. Pavlov [8]

[1] Centre for Cognitive Science, Jagiellonian University, Kraków, Poland
[2] Poznan Laboratory of Affective Neuroscience, Faculty of Psychology and Law, SWPS University, Poznań, Poland
[3] Institute of Psychology, Faculty of Education and Psychology, University of the National Education Commission, Kraków, Poland
[4] Institute of Experimental Psychology, Heinrich Heine University, Düsseldorf, Germany
[5] Doctoral School in the Social Sciences, Jagiellonian University, Kraków, Poland
[6] Department of Education and Counseling Psychology, Santa Clara University, Santa Clara, USA
[7] School of Psychology, University of Leeds, Leeds, United Kingdom
[8] Institute of Medical Psychology and Behavioral Neurobiology, University of Tuebingen, Tuebingen, Germany

***Corresponding author:**
Magdalena Senderecka
Centre for Cognitive Science, Jagiellonian University
Ingardena 3, 30-060 Kraków, Poland
e-mail: magdalena.senderecka@uj.edu.pl

Competing interests:
The authors disclose no conflicts of interest related to this manuscript.

Funding:
#EEGManyLabs is funded by the DFG (PA 4005/1-1) and the UK Biotechnology and Biological Sciences Research Council (BB/X008428/1). This replication is supported by a grant funded by the Strategic Program Excellence Initiative at the Jagiellonian University, Kraków, Poland. The funders had no role in study design or the decision to submit the work for publication.

Abstract

Error monitoring is a key cognitive control process that enables individuals to evaluate their actions and detect failures in goal-directed behavior. A well-established neural correlate of error monitoring is the error-related negativity (ERN), an event-related potential that emerges within 100 ms of an incorrect response. A landmark study by Hajcak et al. (2005) found that ERN amplitude is modulated by the motivational significance of errors. In Experiment 1, errors in high-reward trials elicited larger ERNs than those in low-reward trials. In Experiment 2, ERNs were larger when participants were subject to social evaluation compared to a no-evaluation condition. The present study aims to conduct a high-powered (at least 126 participants after exclusions for both Experiment 1 and 2), multi-lab replication of this research as part of the #EEGManyLabs initiative (Pavlov et al., 2021). We will manipulate motivational context via points-based monetary incentives and social evaluation using a flanker task. We predict more pronounced ERN amplitudes in high- versus low-reward trials, and under social evaluation compared to the control condition. By testing the robustness of these effects, this study will critically assess whether motivational significance consistently modulates early error processing. If successful, the replication would strengthen the claim that the ERN functions as a motivationally tuned signal, integrating error detection with goal-relevant context. If not, it would raise important questions about how motivation interacts with error monitoring. In either case, the study will offer valuable insights into the motivational dynamics of cognitive control, with implications for both theory and experimental research.

Keywords: error-related negativity; error significance; event-related potentials, replication; #EEGManyLabs

1. Introduction

Error monitoring refers to the evaluative process of detecting errors in ongoing actions, a key component of the performance monitoring system (Fu et al., 2023; Kirschner & Ullsperger, 2024). Identifying such deviations enables adaptive adjustments aimed at preventing future mistakes. Errors vary in their affective and motivational significance: some are negligible, as they do not lead to negative social evaluations or other forms of punishment, while others are critical, resulting in negative assessments, loss of rewards, or even threats to survival. Given that mental states and processes are widely believed to have corresponding neural representations–an idea supported by both reductionist accounts in the philosophy of mind (e.g., Churchland, 1986) and empirical findings in cognitive neuroscience (e.g., Poldrack, 2010; Kriegeskorte & Douglas, 2018)–this variability in error significance should be reflected in neural responses.

One of the most commonly studied neurophysiological markers of error detection is the error-related negativity (ERN; Gehring et al., 1993), also referred to as the error negativity (Ne; Falkenstein et al., 1991). The ERN is an event-related potential (ERP) component extracted from electroencephalographic (EEG) recordings that manifests as a sharp negative deflection over the mediofrontal scalp regions within 100 ms after an incorrect motor response. The neural sources of the ERN have traditionally been localized in the medial frontal cortex, most commonly in the dorsal anterior cingulate cortex (dACC) (Brázdil et al., 2005; Debener et al., 2005; Dehaene et al., 1994; Gruendler et al., 2011). However, more recent research suggests that the ERN may also originate from other brain regions, such as the supplementary motor area (SMA) (Bonini et al., 2014) or the pre-supplementary motor area (pre-SMA) (Fu et al., 2019; Iannaccone et al., 2015).

Gehring et al. (1993) were the first to find that the ERN varies with the relative weight participants assign to the accuracy of their actions, in accordance with task instructions. They observed that the ERN amplitude is more pronounced when participants strive for accurate performance, and diminished when response speed is emphasized at the expense of accuracy. This effect was further corroborated by Falkenstein et al. (2000). Following these initial works, subsequent studies began to examine the extent to which the ERN is modulated by motivational and affective factors (for early contributions, see Hajcak et al., 2004; Luu et al., 2000, 2003; Pailing & Segalowitz, 2004). Within this line of research, Hajcak, Moser, Yeung,

and Simons (2005) conducted a seminal study examining whether and how the perceived significance of errors modulates ERN amplitude. To investigate this relationship, they employed two experimental manipulations: points-based incentives and social evaluation.

In their first experiment, the significance of errors was manipulated by associating different point values with each trial in a modified flanker task. Participants were told they could earn either 5 or 100 points for a correct and fast response, with the goal of maximizing their score. Thus, each error signified a loss of potential points. Although participants' actual performance did not affect their final monetary compensation, they were led to believe that the points earned would be converted into bonus money. This made errors on 100-point trials more significant than those on 5-point trials in terms of perceived loss. The results revealed that errors on high-reward trials elicited a more pronounced ERN, suggesting that this neural response is sensitive to the motivational significance of errors.

In the second experiment, error significance was manipulated through social evaluation. Participants completed the study in two conditions: with and without social evaluation. In the social evaluation condition, they were informed that their performance on the flanker task would be monitored by an experimenter sitting next to them and compared to the performance of previous participants. This introduced an element of social judgment, thereby increasing the personal and motivational significance of their actions. In contrast, the control condition involved performing the task without any mention of social evaluation and in the absence of an experimenter. The results showed that errors made in the social evaluation condition elicited larger ERNs compared to the control condition, further highlighting the sensitivity of this neural response to the motivational significance of errors (see Figure 1 and Table 1 for further details on the task designs).

Taken together, these two experiments (Hajcak et al., 2005) demonstrated, first, that the influence of motivational factors on response monitoring is indeed reflected at the neural level in the amplitude of the ERN. Second, the findings showed that the ERN amplitude can be modulated by multiple methods of increasing the motivational significance of errors, whether through points-based monetary incentives or social evaluation. Overall, this study provided evidence from two distinct manipulations that the ERN is sensitive to the personal significance of errors.

Notably, the effect of motivational manipulation in both experiments was limited to error monitoring and was not reflected in the amplitude of other ERP components, such as the correct-response negativity (CRN), associated with the evaluation of successful responses (Vidal et al., 2000, 2003), or the P3 component, linked to the cognitive processing of presented stimuli (Polich, 2007). The inclusion of the CRN in the analysis aimed to determine whether the effect extended to response monitoring more broadly, regardless of response accuracy. Additionally, the P3 component was examined to assess whether the observed effects could be attributed to differences in attentional orienting or arousal. These processes might influence brain activity prior to response onset–just after stimulus presentation–and artificially induce differences in subsequent ERN amplitude due to component overlap. The analyses revealed that the effect of motivational manipulation was selective (within the three components examined) and restricted to early error monitoring, as reflected in ERN amplitude.

The findings from Hajcak et al. (2005), cited over 700 times (Google Scholar, November 2025), have had significant implications for both theoretical accounts of the ERN and subsequent experimental research. Prominent theories regarding the functional significance of the ERN propose that it reflects mechanisms that track the mismatch between intended actions and executed responses, resulting in error detection (Falkenstein et al., 2000; Gehring et al., 1993), or that it signals conflict between competing correct and incorrect response representations (Botvinick et al., 2001; Yeung et al., 2004; for a review, see Larson et al., 2014). Other leading accounts link the ERN to negative reinforcement learning (Holroyd & Coles, 2002) or the updating of expectations regarding action outcomes, as described in the predicted response outcome model (Alexander & Brown, 2010, 2011). Following Hajcak et al. (2005) and related studies (e.g., Chiu & Deldin, 2007; Hajcak et al., 2004; Luu et al., 2000, 2003; Pailing & Segalowitz, 2004), increased attention has been directed toward the role of affective and motivational factors in the mechanisms underlying error detection. Importantly, these studies contributed to the development of a theoretical perspective suggesting that the ERN reflects the affective and motivational significance of errors (Aarts et al., 2013; Proudfit et al., 2013; for a review, see Hajcak, 2012), their aversiveness (Hajcak & Foti, 2008; Jackson et al., 2015), or mechanisms triggering rapid defensive responses (Weinberg et al., 2012, 2015, 2016). According to this approach, the response monitoring system generates negatively valenced affective signals–reflected in the ERN amplitude–when actions deviate from motivational goals, and these signals may be further modulated by contextual factors

that influence the perceived significance of errors. Within this theoretical perspective, the ERN has also been proposed as an endophenotype for the development and maintenance of various psychopathological symptoms, including both internalizing and externalizing mental disorders (Härpfer et al., 2025; Lutz et al., 2021; Manoach & Agam, 2013; Pasion & Barbosa, 2019), with more negative ERN amplitudes observed in obsessive-compulsive disorder (Riesel, 2019; Riesel et al., 2019b), anxiety (Hajcak et al., 2019; Meyer, 2017; Moser et al., 2013; Saunders & Inzlicht, 2020), and depression (Clayson et al., 2020; Moran et al., 2017), and blunted ERN amplitudes in attention-deficit/hyperactivity disorder (Shiels & Hawk, 2010) and substance use disorder (Luijten et al., 2014).

Apart from their theoretical impact, Hajcak et al. (2005) and related works (e.g., Chiu & Deldin, 2007; Hajcak et al., 2004; Luu et al., 2000, 2003; Pailing & Segalowitz, 2004) also sparked a rich line of experimental research, showing that ERN amplitude depends on state-related manipulations involving motivation (for reviews, see Weinberg et al., 2012; Nuñez-Estupiñan et al., 2022). In general, such manipulations influence the degree of personal engagement and the perceived importance of the task, thereby heightening the motivational significance of committed errors (for a review of studies on motivation and cognitive control interactions, see Botvinick & Braver, 2015). Increased ERN magnitude has been observed in studies involving the administration of rewards or encouraging feedback, as well as punishments or derogatory feedback (e.g., Cole et al., 2022; Ganushchak & Schiller, 2008; Leue et al., 2017; Maruo et al., 2016; Potts, 2011; Riesel et al., 2012; Saunders et al., 2015; Stürmer et al., 2011; Unger et al., 2012; Wiswede et al., 2009). More pronounced ERNs have also been reported in adults following the enhancement of perceived autonomy and self-determination (Legault & Inzlicht, 2013), in competitive compared to cooperative contexts (García Alanis et al., 2019), or in situations involving disapproval in a social context (Boksem et al., 2011). Similar effects were observed in adolescents and children when a peer (Kim et al., 2005; Niu et al., 2023) or a critical parent (Meyer et al., 2019) was observing them during the completion of an experimental task.

On the other hand, some similarly designed studies have failed to find a clear association between ERN amplitude and short-term motivational states (e.g., Boksem et al., 2008; Clayson et al., 2012; Elkins-Brown et al., 2016; Holmes & Pizzagalli, 2010), or have found such effects only to a limited extent–for example, only in female participants but not males (Clayson et al., 2025), or when participants' actions were observed by an unsupportive but not

a supportive partner (Palmwood & Simons, 2021). Discrepancies across studies may stem from multiple methodological factors. First, differences in sample size can affect the precision of effect-size estimates. Significant effects observed in small samples (e.g., the 19 participants in Hajcak et al., 2005, Experiment 2) may reflect true effects but are also more susceptible to overestimation due to sampling variability. Second, effects may vary with sample composition. For instance, Wiswede et al. (2009) reported a significant effect of motivation on ERN amplitude in a sample consisting solely of female participants, but this finding was not replicated by Clayson et al. (2012) using the same paradigm in a larger and more gender-diverse sample. Third, divergent findings may result from different task designs, e.g., the ratio of congruent to incongruent trials (Seow et al., 2020). Fourth, differences may also reflect data processing choices that influence both data quality and experimental effects, such as highpass and lowpass filtering, ocular artifact correction, reference, baseline correction, electrodes of interest, ERN component scoring (e.g., Clayson et al., 2021a; Feuerriegel & Bode, 2022; Meyer et al., 2017). Finally, differences may arise from variation in measurement reliability, which strongly depends on the number of error trials available per condition and retained for averaging–a parameter that differs substantially across experiments (Clayson et al., 2021b; Fischer et al., 2017).

Apart from the already mentioned methodological reasons, inconsistency between studies may also stem from stable affective and motivational characteristics of participants. A body of previous work suggests that the influence of experimental manipulations on the ERN may be moderated by individual differences (for a review, see Nuñez-Estupiñan et al., 2022). For instance, in some studies, enhanced ERN amplitudes in punishment compared to control condition were observed only in individuals high in anxiety (Meyer & Gawlowska, 2017; Riesel et al., 2019a) or high in punishment sensitivity (Boksem et al., 2008), as measured with the Behavioral Inhibition System scale (Carver & White, 1994). In other studies, the motivational effect of punishment on ERN amplitude was observed at the group level, however, this main effect was further qualified by a moderating influence of trait anxiety (Riesel et al., 2012; Cole et al., 2022). Additional research has shown more pronounced ERN amplitudes only in participants high in neuroticism and low in conscientiousness on motivated trials with high monetary reward compared to a control condition (Pailing & Segalowitz, 2004), and in individuals high in worry when they were exposed to uncertain (as compared to certain) evaluative feedback (White et al., 2018). In turn, Endrass et al. (2010) reported an ERN amplitude enhancement between standard and punishment conditions only in healthy

controls, but not in patients with obsessive-compulsive disorder, who exhibited comparable ERN amplitudes across both conditions. This discrepancy was interpreted as reflecting patients' inability to down-regulate their error monitoring in response to external demands. Taken together, these findings suggest that the relationship between ERN amplitude and motivation may be more fragile, context-dependent, and less consistent across individuals than previously assumed. Thus, given the presence of both confirmatory and mixed or inconclusive results, further investigation is warranted.

The present study aims to further examine the relationship between ERN amplitude and motivation by conducting a high-powered direct replication of two separate experiments originally reported by Hajcak et al. (2005). By replicating these experiments, we will test whether the ERN functions as a motivationally tuned signal across two distinct goal-relevant contexts, created by points-based monetary incentives and social evaluation. Replication success will be evaluated independently for each experiment. If both experiments are successfully replicated, the results would provide support for the theoretical account linking ERN amplitude to error significance across different motivational contexts. If only one experiment is replicated, this will help identify the specific context in which motivational modulation of ERN amplitude occurs. If neither effect is observed, the results may indicate that: (1) the influence of motivational states on the ERN is smaller than previously assumed and not detectable even in a large sample; (2) this influence varies across individuals and depends on sample composition or individual traits; and/or (3) the tasks used may not always be sufficient to ensure sensitivity of the ERN to the studied phenomena (e.g., the number of errors in the flanker task with equiprobable congruent and incongruent trials may be insufficient for reliable ERN estimation).

Our replication will be carried out through a multi-lab collaboration within the large international initiative #EEGManyLabs (for details, see Pavlov et al., 2021). We will employ experimental designs and analysis pipelines as closely aligned with the original study as possible. Following Hajcak et al. (2005), we will test two independent hypotheses that the ERN amplitude is larger (more negative): (1) when higher points are awarded for correct responses compared to lower points; and (2) when performance is subject to social evaluation compared to when it is not. Ultimately, we should be able to determine whether, and how consistently, variability in error significance is reflected in ERN amplitude.

2. Method

The present manuscript is a Stage 1 Registered Report and adheres to the recommended open science practices for psychophysiological research (Garrett-Ruffin et al., 2021). A project repository is hosted on the Open Science Framework (OSF; https://osf.io/un3st/) and will serve as a central hub, linking to raw EEG data, as well as data processing and analysis code.

2.1. Participants

Participants will be recruited either from the general community via local advertisements or from institutional participant pools. Since the original study was conducted with young adult participants, the age range will be set between 18 and 35 years to mitigate potential concerns regarding age-related cognitive decline. All participants will be healthy, free of medications, with no history of neurological or psychiatric disorders, and will have normal or corrected-to-normal vision as confirmed by self-declaration. Sex will be included as a factor in exploratory follow-up analyses; however, sex balance will not be controlled during recruitment in order to remain consistent with the original study. The sample is anticipated to be composed largely of individuals from W.E.I.R.D. (Western, Educated, Industrialised, Rich, and Democratic) backgrounds. Replicating labs located in the same city will cooperate during participant recruitment to ensure the study is run on distinct samples and avoid collecting data from the same participant twice. Prior to the study, participants will provide written informed consent to take part in the study (Experiment 1 or Experiment 2) and to the posting of their anonymized raw data. They will be compensated for their time either financially, according to local policies, or by receiving course credits. Additionally, participants in Experiment 1 will be informed that they can earn a monetary bonus (in addition to regular remuneration or course credits) based on their performance. However, regardless of their response speed and accuracy, the payout will be the same for all participants. The reward amount will be determined based on local standards.

2.2. Sample size and power analysis

In the original study, 22 undergraduate students participated in Experiment 1 (two were excluded due to near-perfect task performance), and 19 undergraduate students participated in Experiment 2 (one was excluded due to a technical malfunction). To determine the minimum target sample size for our replication study, we conducted power analyses using G*Power 3 (Faul et al., 2007). In line with the #EEGManyLabs protocol (Pavlov et al., 2021), power was

set at 0.90 with $\alpha = .02$ (see also Camerer et al., 2018; Lewis et al., 2020; Schäfer & Schwarz, 2019). Analyses were specified as one-tailed, reflecting the strong a priori directional hypothesis that increased motivational significance is associated with more pronounced ERN amplitudes. The effect sizes were based on those reported by Hajcak et al. (2005) for the significant main effect of trial value on ERN amplitude in Experiment 1, $F(1,19) = 7.53$, $p < .05$; and the significant main effect of the experimental condition on ERN amplitude in Experiment 2, $F(1,17) = 6.44$, $p < .05$. These two effects were chosen as they relate to the main research questions of the study. For a two-level within-subject factor, these correspond to paired-samples statistics of $t = 2.74$ and $t = 2.54$, yielding Cohen's $d = 0.61$ and $d = 0.60$, respectively. For completeness, the corresponding repeated-measures correlations are $r = 0.53$ and $r = 0.52$, respectively. For the present power analysis, we conservatively targeted half of these effect sizes ($d = 0.30$), as evidence suggests that effect sizes in replication studies tend to be about half the size of those in original studies (Open Science Collaboration, 2015), similar to how effect sizes in preregistered studies are typically smaller than in studies without preregistration (Schäfer & Schwarz, 2019). Power analyses (paired-samples *t*, one-tailed, $\alpha = .02$, $1 - \beta = .90$) indicated that $d = 0.30$ requires 126 participants (after exclusions) for both Experiment 1 and Experiment 2. Power was defined for the pooled analysis across laboratories. Data collection will proceed for 1.5 years following Registered Report Stage 1 acceptance, or until a maximum of $N = 216$ eligible participants (after exclusions, corresponding to 99% power for the same target effect, $d = 0.30$) has been reached (whichever occurs first). If the 1.5-year limit is reached and fewer than $N = 126$ eligible participants are available, data collection will continue only until $N = 126$ is reached. Throughout data collection, datasets will be processed solely for quality control and exclusion purposes, and no outcome-relevant analyses or inspection of the critical dependent variables will be conducted prior to the formal completion of data collection. The additional datasets (if available) would serve to increase precision and statistical power for the main hypothesis tests, and support a more reliable characterization of between-site variability and related exploratory analyses. No interim hypothesis testing will be performed before the stopping rule has been met. Experiments 1 and 2 will be conducted on two separately recruited samples.

At the time of the Stage 1 submission, four replicating labs had confirmed their commitment to data collection (see Table 2 for details). Data will be collected from labs in Kraków, Poland (with two replicating labs), Poznań, Poland, Düsseldorf, Germany, and other labs will be

invited to join the project at Stage 2. Individual labs will collect data according to their local capacity.

2.3. Overall procedure

Replicating labs will obtain approval from their local institutional review boards or ethics committees to conduct the study and share de-identified data. The procedures outlined in Hajcak et al. (2005) will serve as a model for the present replication. Participants will receive consistent information across replicating labs regarding the general purpose and course of the study and will provide written informed consent. After completing the study, they will be debriefed about its actual aims and procedures. Instructions will be adapted to the local language. EEG signals will be recorded during both experiments.

In Experiment 1, participants will perform a modified flanker task. They will be informed that each trial will begin with the presentation of a '5' or '100' cue, indicating the value of the trial. The value will refer to the points they could earn for an accurate and fast response. Participants will be encouraged to maximize their point earnings and led to believe that these points will be converted into money at the end of the study. In reality, their remuneration will be unrelated to performance. Participants will receive a fixed monetary reward for their performance, determined based on local policies, along with standard compensation for their time spent in the laboratory, either as financial payment or course credits. They will also be debriefed about the actual aim and procedure of the study, including the use and necessity of deception. Specifically, the debriefing will explain that participants were led to believe their performance would influence their compensation in order to induce genuine motivation and engagement with the task, which was essential for addressing the study's research questions.

In Experiment 2, participants will perform a standard flanker task twice – once in the social evaluation condition and once in the control condition – with a short break in between. The order of conditions will be counterbalanced across participants. In the social evaluation condition, participants will be informed that their speed and accuracy will be evaluated in real time by a research assistant and compared to the performance of previous participants. They will also be told that they will receive feedback on their relative performance at the end of the study. During the task, a male research assistant will be seated approximately 0.5 meters to the side of the participant at a desk with a laptop, supposedly receiving a signal from their computer. In reality, no actual evaluation will occur. At the end of the experiment, all

participants will be notified that they performed above average. In turn, in the control condition, participants will be informed that their responses will not be evaluated and instructed to simply respond as quickly and accurately as possible. The research assistant will not be present in the room. If the control condition occurs second in the counterbalanced order, participants will be explicitly reminded that no formal performance evaluation will take place. After completing the entire study, participants will be compensated for their time spent in the laboratory. They will also be debriefed about the nature of the experimental manipulation, including the use and rationale for deception. Specifically, the debriefing will explain that participants were led to believe they were being evaluated in order to create a realistic sense of social evaluative pressure, which was critical for examining the effects of perceived evaluation on task performance.

Some of the labs involved in this replication will additionally collect 8 minutes of resting-state EEG data as part of the #EEGManyLabs Resting State spin-off project (https://osf.io/sp3ck/). Participants will also complete a series of self-report scales (available in local language versions) assessing affective, motivational, and personality-related traits, as well as handedness and sleepiness. These scales include: (1) the Apathy-Motivation Index (AMI; Ang et al., 2017); (2) the Behavioral Inhibition System/Behavioral Activation System Scales (BIS/BAS; Carver & White, 1994); (3) the Big Five Inventory – Short Version (BFI–S; Gerlitz & Schupp, 2005); (4) the Center for Epidemiologic Studies Depression Scale (Radloff, 1977); (5) the Edinburgh Handedness Inventory (EHI; Oldfield, 1971); (6) the Global Motivation Scale (GMS; Guay et al., 2003); (7) the Intolerance of Uncertainty Scale – Short Version (IUS-12; Carleton et al., 2007); (8) the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990); (9) the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990); (10) the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988); (11) the Positive Valence Systems Scale (PVSS-21; Khazanov et al., 2020); (12) the Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ; Torrubia et al., 2001); and (13) the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1970). Data from brief state-related questionnaires will be collected immediately after the EEG resting-state recording, while the remaining trait-based questionnaires will be completed either after the entire study session (in-lab) or in advance of the participant's lab visit (remotely). These additional self-report data may be included in exploratory analyses (e.g., examining whether individual differences modulate the ERN–motivation relationship) and will also serve as a foundation for future research and publications.

2.4. Flanker tasks

In Experiment 1, participants will complete a modified version of the flanker task (Eriksen & Eriksen, 1974), based on the original study (see Figure 1, panel A). During the task, they will be presented with sets of five arrowheads (‘‘<<<<<’’, ‘‘<<><<’’, ‘‘>>>>>’’, or ‘‘>><>>’’) and instructed to press the left or right mouse button according to the direction of the center arrowhead. These four imperative stimuli will be presented in random order, with half being congruent (‘‘<<<<<’’ or ‘‘>>>>>’’) and the other half incongruent (‘‘<<><<’’ or ‘‘>><>>’’).

Each stimulus will be preceded by a randomly presented ‘5’ or ‘100’ cue, indicating the point value of the trial. The cue will remain on the screen for 2000 ms. The imperative stimulus will appear 200 to 800 ms after cue offset and will remain visible for 200 ms. The interval between the offset of the imperative stimulus and the onset of the next cue will vary between 1700 and 2300 ms. A fixation mark (+) will be displayed between all cues and stimuli. Each set of arrowheads will occupy 1.31° of visual angle vertically and 9.21° horizontally at a viewing distance of approximately 65 cm. All cues, stimuli and fixation marks will be presented in white font against a black background and positioned in the center of the screen.

Participants will first complete two blocks of 48 practice trials. The main experiment will consist of 12 blocks of 48 trials (576 trials in total), with congruent and incongruent trials randomly distributed and fully crossed with the ‘5’ and ‘100’ point value cues (i.e., congruent and incongruent trials will occur equally often with both cue values, in random order). Each block will be initiated at the participant's own pace. In line with the original study, participants will be instructed to respond both correctly and as quickly as possible. They will be informed that points are awarded only for responses that are both fast and accurate, so both conditions must be met simultaneously. Task design details, beyond those reported in the original manuscript, were provided by the Authors of the original study.

In Experiment 2, participants will perform a standard arrowhead version of the flanker task, also based on the original study (see Figure 1, panel B). The design of the task will be identical to that described for Experiment 1, with two exceptions. First, there will be no ‘5’ or ‘100’ cue, so imperative stimuli will be presented one after the other, in random order, with an inter-trial interval lasting between 1700 and 2300 ms. Second, the task will be performed twice: once in the social evaluation condition and once in the control condition (in

counterbalanced order). Participants will be instructed to respond both quickly and accurately. They will complete 12 blocks of 48 trials (576 trials) in each of the two experimental conditions. Thus, the total number of trials for the entire experiment will be 1152.

To remain consistent with the original study, participants will not receive any feedback regarding their performance in either task. A detailed overview of the flanker task parameters is presented in Table 1. The original code for both flanker tasks is no longer available; therefore, it has been rewritten in PsychoPy (Peirce, 2007) and can be adapted for use with other task programming software. The PsychoPy code is accessible on GitHub: https://github.com/filyp/hajcak2005_replication.

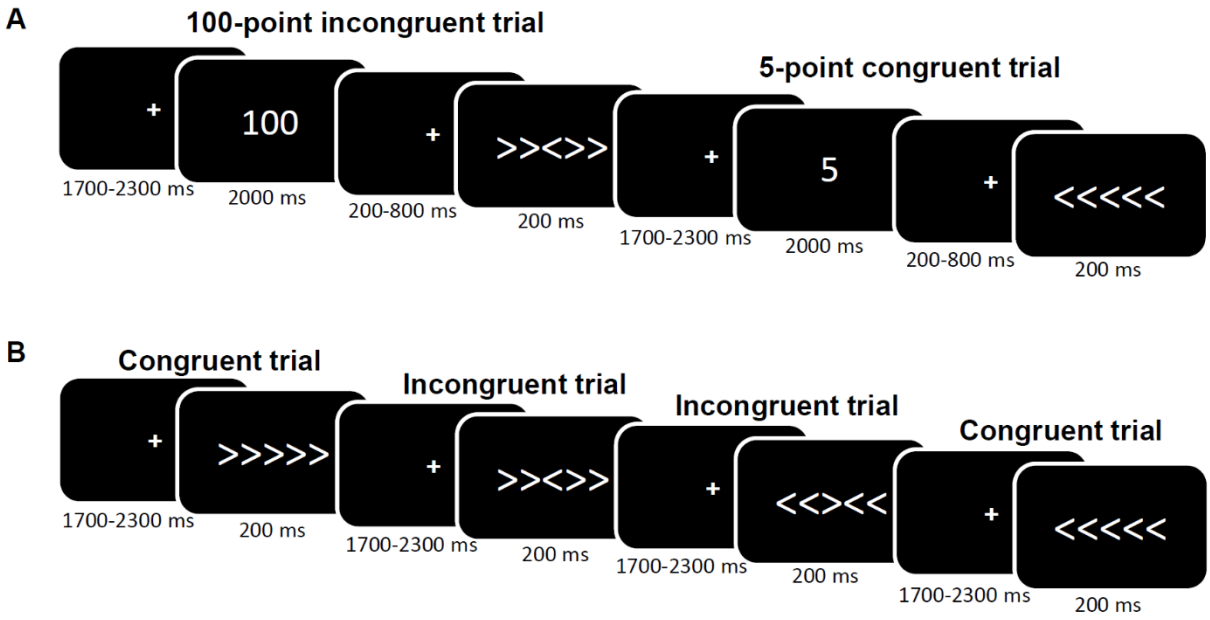

Fig. 1 – Flanker task designs. Panel A illustrates the modified arrowhead version of the flanker task from Experiment 1, with a '5' or '100' cue indicating the trial value. Panel B depicts the standard arrowhead version of the flanker task from Experiment 2. Congruent and incongruent trials are equiprobable in both experiments.

Table 1 – Flanker task parameters in the original study and the direct replication

Parameter	Hajcak et al. (2005)	#EEGManyLabs direct replication
Stimulus set & mapping	Five-arrow flanker arrays (<<<<<, <<><<, >>>>>, >><>>); respond to central arrow; 50% congruent/incongruent	Identical stimulus set and mapping
Viewing distance & visual angle	~65 cm; 1.31° (vertical) × 9.21° (horizontal); white on black, centrally presented	Same parameters
Fixation	'+' presented before each stimulus	Identical fixation procedure
Cue (Exp. 1 only)	Digit '5' or '100' shown for 2000 ms before each trial	Identical cue presentation
SOA (cue → stimulus, Exp. 1 only)	200–800 ms (variable) after cue offset	Identical SOA
Stimulus duration	200 ms	Identical duration
ITI	1700–2300 ms (variable), in Exp. 2, same interval without cue	Identical ITI
Practice trials	2 blocks × 48 trials	Identical number of blocks and trials
Main task – Exp. 1 (points)	12 blocks × 48 trials = 576 trials total	Identical structure
Main task – Exp. 2 (social evaluation)	Two full task runs of 12 blocks × 48 trials each (evaluation and control), 1152 trials total; order counterbalanced; assistant seated nearby (~0.5 m) during evaluation condition	Identical structure
Randomization	Stimuli presented in random order; in Exp. 1, congruent and incongruent trials fully crossed with the '5' and '100' cues	Identical randomization scheme
Instructions	Speed and accuracy equally emphasized; in Exp. 1, points for fast and accurate responses converted into money (deceptive incentive)	Same instructions (translated and harmonized across labs)
Feedback	No feedback between trials or blocks	Identical structure
Devices	Desktop PC, standard computer mouse responses (right hand, index and middle fingers)	Equivalent setup

ITI - inter-trial interval; SOA - stimulus onset asynchrony

2.5. EEG recording and preprocessing

Detailed information about each system and its recording parameters is presented in Table 2.

Table 2 - Overview of EEG systems and recording parameters at each replicating lab

Participating university	Amplifier system	Electrode types / channel number	Reference / Ground channel location	Hardware filters	Sampling rate	Offline reference	Stimulus presentation language	Laboratory environment	Expected number of participants (after exclusions) in Experiment 1 / 2
Jagiellonian University, Poland	BioSemi Active-Two (BioSemi, Amsterdam, Netherlands)	Ag/AgCl active electrodes / 64 scalp channels, 4 EOG channels	CMS/DRL	HP: 0 Hz LP: Auto-adjusted (~1/5 sampling rate)	512 Hz	average of 2 reference channels (mastoids)	Psychopy, Polish	dimly lit, soundproofed	36 /52
University of the National Education Commission, Poland	BioSemi Active-Two (BioSemi, Amsterdam, Netherlands)	Ag/AgCl active electrodes / 64 scalp channels, 4 EOG channels	CMS/DRL	HP: 0 Hz LP: Auto-adjusted (~1/5 sampling rate)	512Hz	average of 2 reference channels (mastoids)	Psychopy, Polish	dimly lit, soundproofed	30 / 37
SWPS University, Poland	ActiCHamp (Brain Products, Gilching, Germany)	Ag/AgCl active electrodes, 64 scalp channels, 2 EOG channels	CPz, FPz	HP: 0 Hz LP: Auto-adjusted (~1/4 sampling rate)	500 Hz	average of TP9 and TP10 from the 64-channel scalp electrode set	Psychopy or Presentation, Polish	dimly lit, soundproofed, electrically shielded	30 / 37
University Düsseldorf, Germany	ActiCHamp (Brain Products, Gilching, Germany)	Ag/AgCl active electrodes, 64 scalp channels, 2 EOG channels	CPz, FPz	HP: 0 Hz LP: Auto-adjusted (~1/4 sampling rate)	500 Hz	average of TP9 and TP10 from the 64-channel scalp electrode set	Psychopy or Presentation, German	dimly lit, soundproofed, electrically shielded	30 / 0

EOG = electrooculogram channel; HP = high pass; LP = low pass

The EEG signal will be continuously digitized at a sampling rate of at least 500/512 Hz, depending on the setup. After recording, all EEG data will be processed using a pipeline summarized in Table 3, which closely mirrors the procedures described in the original study by Hajcak et al. (2005). Continuous EEG data will be down-sampled to 250 or 256 Hz to approximate the original study's sampling rate of 200 Hz. The signal will be re-referenced offline to the average of both mastoid electrodes, or alternatively, to the near-mastoid electrodes TP9 and TP10. Offline filtering will be applied using an infinite impulse response (IIR) Butterworth filter with a low cutoff frequency of 0.05 Hz (24 dB/octave slope), a high cutoff frequency of 35 Hz (12 dB/octave slope), and a notch filter at 50 Hz or 60 Hz,

depending on the local power-line frequency. Note that in the original study, a 0.05 to 35 Hz bandpass filter was applied online. However, since replicating labs will use online filtering settings consistent with their own hardware, the bandpass filter will be applied offline.

Response-locked epochs will be extracted using a time window from 200 ms before to 600 ms after the button press, consistent with the ERP visualization in the original study. Eye blinks and saccadic eye movements will be corrected using the Gratton, Coles, and Donchin method (Gratton et al., 1983). Trials with a 'flat' signal (at least 25 ms of invariant data on any channel) or excessive physiological artifacts (voltage exceeding ±100 μV) will be rejected. Note that in the original study, data segments were excluded based on invariant data and/or falling out of the range of the analog-to-digital converter. Since all replicating labs will record EEG signals using different setups than the original study, the rejection criteria described above will be used instead (see Paul et al., 2025, for a similar solution in an analogous replication study from the #EEGManyLabs initiative). Trials with response times below 200 ms or above 800 ms will also be rejected, consistent with the procedure used in the original study. Response-locked ERPs will be baseline adjusted using a 100-ms window immediately preceding the participant's button press (-100 to 0 ms). ERN activity will be scored as the mean amplitude between 0 and 100 ms following the participant's response from the averaged signal of electrodes Fz and Cz, where the ERN was most pronounced in the original study. Note that Fz and Cz were two of the three electrode sites (Fz, Cz, Pz) included in Hajcak et al. (2005). However, because running separate analyses for three different sites would increase the risk of obtaining partially significant effects, making it difficult to determine replication success unambiguously, we decided to restrict the key analyses to the averaged Fz–Cz signal. All EEG recording and preprocessing procedures will be identical in Experiment 1 and Experiment 2. Data preprocessing will be performed by the lead lab. The code prepared for this purpose is available online (https://osf.io/un3st/). All replicating labs will regularly monitor their data collection to track participant exclusions (based on predefined criteria described in the *Participant- and trial-level exclusion criteria* section below) and ensure that the required number of eligible participants per lab is achieved. Cases of exclusion will be additionally verified and confirmed by the lead lab.

Table 3 - Preprocessing pipelines in the original study and the direct replication

Offline preprocessing step	Hajcak et al. (2005)	#EEGManyLabs direct replication
Downsampling	Not applicable (signal recorded with a sampling rate of 200 Hz)	Downsampled to 250/256 Hz
Offline re-referencing	Average of both mastoid electrodes	Average of both mastoid electrodes, or near-mastoid electrodes TP9 and TP10
Offline filtering	Not applicable (a 0.05–35 Hz bandpass filter applied online)	Infinite impulse response Butterworth filter with: (1) 0.05 Hz low cutoff (slope: 24 dB/oct); (2) 35 Hz high cutoff (slope: 12 dB/oct); (3) notch filter at 50 or 60 Hz (depending on local power-line frequency).
Segmentation	Response-locked: -200 to 600 ms (as inferred from ERP visualizations)	Response-locked: -200 to 600 ms
Ocular artifact correction	Gratton, Coles, and Donchin's algorithm (Gratton et al., 1983)	Gratton, Coles, and Donchin's algorithm (Gratton et al., 1983)
Artifact rejection	Trials with 'flat' signal (≥ 25 ms of invariant data on any channel) or excessive physiological artifacts (signal beyond A/D converter range)	Trials with 'flat' signal (≥ 25 ms of invariant data on any channel) or excessive physiological artifacts (voltage exceeding ±100 μV)
Trial rejection based on RT	RT < 200 or RT > 800	RT < 200 or RT > 800
Baseline correction	Response-locked: -100 to 0 ms	Response-locked: -100 to 0 ms
ERN scoring	Mean amplitude between 0–100 ms post-response at Fz, Cz, and Pz	Mean amplitude between 0–100 ms post-response at averaged Fz and Cz (selected as the most appropriate measure for determining replication success)

A/D = analog-to-digital; RT = response time

2.6. Data analysis

2.6.1. Participant- and trial-level exclusion criteria

Data-screening and exclusion procedures will follow those applied in the original study, with additional criteria introduced to ensure high data quality and interpretability of the EEG signal. In the original study, participants were excluded if they committed fewer than five errors in either condition of interest, and individual trials with response times outside the predefined range (200–800 ms) were removed. In the present replication, we will adopt these criteria and further extend the exclusion pipeline to include checks for the error-to-correct response ratio, EEG signal quality, and the number of valid error trials available at the electrodes of interest after preprocessing and response time filtering.

Specifically, participants will be required to have at least five valid error trials per condition, with the proportion of error trials not exceeding 30% of all responses. The upper limit is intended to eliminate cases where participants' performance might indicate insufficient engagement, random responding, or misunderstanding of task instructions – factors that could distort the neural processes underlying error monitoring. The predefined error rate thresholds should also help maintain comparable trial distributions across participants and labs, thereby reducing the likelihood of systematic differences that could affect interpretation. In addition, participants will be excluded if more than 50% of trials at the electrodes of interest (Fz and Cz) are rejected due to artifacts. These steps will ensure that the retained data represent reliable performance and signal quality. A detailed overview of the exclusion pipeline is presented in Table 4.

Table 4 - Outlier and data exclusion pipeline

Pipeline step	Criterion checked
Step 1: Minimum number of error trials per condition (in line with the original study)	For each participant, check the number of errors made in: (1) the 5-point and 100-point conditions in Experiment 1; or (2) the social evaluation and control conditions in Experiment 2. Participants must have at least five error trials per condition to proceed to the next step.
Step 2: Error-to-correct response ratio	Check whether the number of errors does not exceed 30% of all responses per condition. Specifically, it should be: (1) equal to or less than 87 in the 5-point and 100-point conditions in Experiment 1 (with 288 trials per condition in total); or (2) equal to or less than 173 in the social evaluation and control conditions in Experiment 2 (with 576 trials per condition in total). If the number of errors constitutes 30% or less of all responses per condition, proceed to the next step.
Step 3: Response time filtering (in line with the original study)	Exclude all trials with response times below 200 ms or above 800 ms. After filtering, verify the number of remaining error trials per condition. Participants must still have at least five error trials per condition to proceed.
Step 4: Artifact rejection at electrodes of interest	Perform EEG preprocessing and calculate the percentage of trials rejected due to artifacts at electrodes Fz and Cz for each condition. If less than 50% of trials are rejected for a given condition, proceed to the next step.
Step 5: Final inclusion check	After artifact rejection, confirm the number of remaining trials at Fz and Cz per condition. Participants who still have at least five valid error trials per condition are included in the final analyses.

2.6.2. Check of motivation effects on ERN

The primary effects of interest from the original study are that the ERN amplitude is more pronounced (i.e., more negative) when high points (compared to low points) are associated with correct responses in Experiment 1, and when performance is subject to social evaluation (compared to no evaluation) in Experiment 2. These effects will be tested via planned contrasts comparing the mean ERN amplitude between the two relevant conditions. We will conduct paired-samples *t*-tests on ERN amplitudes derived from the averaged Fz–Cz signal for the two trial-value conditions ('5'-point cue vs. '100'-point cue) in Experiment 1, and analogous paired-samples *t*-tests for the two condition types (social evaluation vs. control) in Experiment 2. We preregister the analyses to be conducted on the EEG signal averaged across Fz and Cz, as the ERN was most pronounced at these sites in the original study. Analyses will be conducted separately for each participating lab. We will report Cohen's *d* for paired samples with 95% confidence intervals. Bayes factors (BF_{10}) will be reported alongside frequentist statistics to provide a complementary assessment of the strength of evidence.

We expect statistically significant main effects of trial value (Experiment 1) and experimental condition (Experiment 2), with numerically larger (i.e., more negative) ERN amplitudes at electrode Fz in high-point compared to low-point trials, and in the social evaluation condition compared to the control condition. A *p*-value of .02 will be used as the threshold for statistical significance, consistent with the #EEGManyLabs protocol (Pavlov et al., 2021) and the author guidelines for *Cortex.* Since the two experiments test independent hypotheses, no alpha correction will be applied across experiments (Rubin, 2021). By preregistering our analyses, our goal is to ensure that the effect size estimated from our sample is reported without bias and irrespective of its statistical significance.

2.6.3. Determination of replication success

Effect sizes (Cohen's *d*) will be extracted from each replicating lab for both experiments based on the planned paired *t*-tests. These effect sizes will be entered into a random-effects meta-analysis (Riley et al., 2010), treating replicating labs as a random effect–provided the number of contributing labs is sufficient–and using the REML estimator for random-effects variance. Separate analyses will be conducted for Experiment 1 and Experiment 2. This random-effects approach addresses, and helps estimate, the effect of heterogeneity in EEG devices and participant samples across replicating labs. We will report the distribution of the

weighted effect sizes, their 95% confidence intervals, and heterogeneity indices. The effect sizes will be presented and illustrated in forest plots. We will use the metafor package (Viechtbauer, 2010) in R for the meta-analysis.

Similarly, we will conduct Bayesian random-effects meta-analyses using the RoBMA package for R (Maier et al., 2023) to estimate the pooled effect size and between-study heterogeneity. Evidence for the successful replication will be evaluated by examining whether the pooled posterior distribution supports an effect in the predicted direction and whether the Bayes factor in favor of the alternative hypothesis (BF_{10}) exceeds 3, which is consistent with moderate-to-strong evidence. We will report the pooled effect (μ), posterior distribution, 98% confidence intervals, and heterogeneity ($\tau 2$).

In accordance with the #EEGManyLabs protocol (Pavlov et al., 2021), replication will be considered successful, separately for each experiment, if a statistically significant meta-analytic estimate ($p < .02$) in the frequentist meta-analysis is observed across replicating labs and the effect is in the expected direction. Frequentist significance alone will not suffice. If the Bayesian meta-analysis provides evidence for the null ($BF_{10} < 0.333$), we will not conclude that the replication is successful. This joint criterion directly addresses the risk that very large samples with a strict alpha could produce small *p*-values that conflict with Bayesian evidence – Lindley's paradox (Lindley, 1957).

2.6.4. Exploratory control analyses

Since several factors cannot be fully controlled within the strict framework of a direct replication, they will be addressed in follow-up exploratory analyses. These analyses are additional and complementary to the preregistered planned contrast, which constitutes our primary confirmatory test. They do not modify the primary hypothesis or the alpha level (α = .02) and will not be considered when determining replication success.

First, because the number of errors tends to correlate with ERN amplitude (Fischer et al., 2017), differences in response accuracy may confound the interpretation of ERN modulation. We will therefore compare mean accuracy between experimental conditions. If a significant difference is observed, accuracy will be entered as a covariate in exploratory models to assess whether the condition effect on ERN amplitude (if observed) remains after controlling for this variable. Second, differences in response times between conditions could lead to differential temporal overlap between stimulus-locked and response-locked ERP components, potentially

affecting the observed ERN amplitude. Such response time differences may arise from distinct response strategies or from variability in the proportion of congruent and incongruent trials (typically associated with faster and slower responses, respectively). We will therefore compare mean response times for error trials and examine whether congruency effects differ between conditions. If significant response time differences are found, these factors will also be considered in exploratory covariate analyses to examine whether the condition-related ERN effects remain after accounting for potential temporal overlap. Third, the original threshold of five valid error trials may be relatively lenient and could result in unstable ERN estimates for some participants (Pontifex et al., 2010). In exploratory follow-up analyses, we will therefore examine whether applying a more conservative threshold (e.g., ≥ 8 error trials) affects the results. Fourth, since previous studies suggest that sex may modulate the relationship between ERN and motivational manipulations, it will be included in separate exploratory analyses, similarly to other individual traits that will be measured in the study. These investigations may also serve as the basis for future publications beyond the scope of the present replication. Importantly, all analyses described above will be considered secondary and will not alter the primary confirmatory test or the criteria for replication success.

Acknowledgements

The #EEGManyLabs core team and lead lab would like to thank the Authors of the original paper for generously sharing critical information that will help ensure the fidelity and quality of this replication attempt.

References

Aarts, K., De Houwer, J., & Pourtois, G. (2013). Erroneous and correct actions have a different affective valence: Evidence from ERPs. *Emotion, 13*(5), 960–973. https://doi.org/10.1037/a0032808

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*(1-2), 29–37. https://doi.org/10.3109/00207459008994241

Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science*, *2*(4), 658–677. https://doi.org/10.1111/j.1756-8765.2010.01085.x

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344. https://doi.org/10.1038/nn.2921

Ang, Y. S., Lockwood, P., Apps, M. A., Muhammed, K., & Husain, M. (2017). Distinct subtypes of apathy revealed by the apathy motivation index. *PloS One*, *12*(1), e0169938. https://doi.org/10.1371/journal.pone.0169938

Boksem, M. A. S., Ruys, K. I., & Aarts, H. (2011). Facing disapproval: Performance monitoring in a social context. *Social Neuroscience*, *6*(4), 360–368. https://doi.org/10.1080/17470919.2011.556813

Boksem, M. A. S., Tops, M., Kostermans, E., & De Cremer, D. (2008). Sensitivity to punishment and reward omission: Evidence from error-related ERP components. *Biological Psychology*, *79*(2), 185–192. https://doi.org/10.1016/j.biopsycho.2008.04.010

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P., & Vidal, F. (2014). Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science*, *343*(6173), 888–891. https://doi.org/10.1126/science.1247412

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, *66*(1), 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Brázdil, M., Roman, R., Daniel, P., & Rektor, I. (2005). Intracerebral error-related negativity in a simple go/nogo task. *Journal of Psychophysiology*, *19*(4), 244–255. https://doi.org/10.1027/0269-8803.19.4.244

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders, 21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Chiu, P. H., & Deldin, P. J. (2007). Neural evidence for enhanced error detection in major depressive disorder. *American Journal of Psychiatry, 164*(4), 608–616. https://doi.org/10.1176/ajp.2007.164.4.608

Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT Press.

Clayson, P. E., Baldwin, S. A., Rocha, H. A., & Larson, M. J. (2021a). The data-processing multiverse of event-related potentials (ERPs): A roadmap for the optimization and standardization of ERP processing and reduction pipelines. *NeuroImage, 245*, 118712. https://doi.org/10.1016/j.neuroimage.2021.118712

Clayson, P. E., Brush, C. J., & Hajcak, G. (2021b). Data quality and reliability metrics for event-related potentials (ERPs): The utility of subject-level reliability. *International Journal of Psychophysiology, 165*, 121–136. https://doi.org/10.1016/j.ijpsycho.2021.04.004

Clayson, P. E., Carbine, K. A., & Larson, M. J. (2020). A registered report of error-related negativity and reward positivity as biomarkers of depression: P-Curving the evidence. *International Journal of Psychophysiology, 150*, 50–72. https://doi.org/10.1016/j.ijpsycho.2020.01.005

Clayson, P. E., Clawson, A., & Larson, M. J. (2012). The effects of induced state negative affect on performance monitoring processes. *Social Cognitive and Affective Neuroscience, 7*(6), 677–688. https://doi.org/10.1093/scan/nsr040

Clayson, P. E., Varkala, K., Baldwin, S. A., Steffen, P. R., Sandberg, J. G., & Larson, M. J. (2025). Spouse support and stress: Gender differences in neural measures of performance monitoring under observation of a spouse. *Social Cognitive and Affective Neuroscience*, nsaf053. https://doi.org/10.1093/scan/nsaf053

Cole, S. L., Cibrian, E., Mirzadegan, I. A., & Meyer, A. (2022). The impact of punishment on error-related brain activity in children. *Developmental Psychobiology, 64*(7), e22318. https://doi.org/10.1002/dev.22318

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*, 11730–11737. https://doi.org/10.1523/JNEUROSCI.3286-05.2005

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*, 303–305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D. J., Buttrick, N. R., Chartier, C. R., ... & Szecsi, P. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science, 3*(3), 309-331. https://doi.org/10.1177/2515245920958687

Elkins-Brown, N., Saunders, B., & Inzlicht, M. (2016). Error-related electromyographic activity over the corrugator supercilii is associated with neural performance monitoring. *Psychophysiology, 53*(2), 159–170. https://doi.org/10.1111/psyp.12556

Endrass, T., Schuermann, B., Kaufmann, C., Spielberg, R., Kniesche, R., & Kathmann, N. (2010). Performance monitoring and error significance in patients with obsessive-compulsive disorder. *Biological Psychology, 84*(2), 257–263. https://doi.org/10.1016/j.biopsycho.2010.02.002

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components II: Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51*, 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feuerriegel, D., & Bode, S. (2022). Bring a map when exploring the ERP data processing multiverse: A commentary on Clayson et al. 2021. *NeuroImage, 259*, 119443. https://doi.org/10.1016/j.neuroimage.2022.119443

Fischer, A. G., Klein, T. A., & Ullsperger, M. (2017). Comparing the error-related negativity across groups: The impact of error-and trial-number differences. *Psychophysiology, 54*(7), 998–1009. https://doi.org/10.1111/psyp.12863

Fu, Z., Sajad, A., Errington, S. P., Schall, J. D., & Rutishauser, U. (2023). Neurophysiological mechanisms of error monitoring in human and non-human primates. *Nature Reviews Neuroscience, 24*(3), 153–172. https://doi.org/10.1038/s41583-022-00670-w

Fu, Z., Wu, D.-A. J., Ross, I., Chung, J. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2019). Single-neuron correlates of error monitoring and post-error adjustments in human medial frontal cortex. *Neuron, 101*(1), 165–177. https://doi.org/10.1016/j.neuron.2018.11.016

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage, 39*(1), 395–405. https://doi.org/10.1016/j.neuroimage.2007.09.001

García Alanis, J. C., Baker, T. E., Peper, M., & Chavanon, M. L. (2019). Social context effects on error-related brain activity are dependent on interpersonal and achievement-related traits. *Scientific Reports, 9*(1), 1728. https://doi.org/10.1038/s41598-018-38417-2

Garrett-Ruffin, S., Hindash, A. C., Kaczkurkin, A. N., Mears, R. P., Morales, S., Paul, K., Pavlov, Y. G., & Keil, A. (2021). Open science in psychophysiology: An overview of challenges and emerging solutions. *International Journal of Psychophysiology, 162*, 69–78. https://doi.org/10.1016/j.ijpsycho.2021.02.005

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gerlitz, J. Y., & Schupp, J. (2005). Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP. *DIW Research Notes, 4*(10.1016), 1–44.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology, 55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Gruendler, T. O. J., Ullsperger, M., & Huster, R. J. (2011). Event-related potential correlates of performance-monitoring in a lateralized time-estimation task. *PloS One, 6*(10), e25591. https://doi.org/10.1371/journal.pone.0025591

Guay, F., Mageau, G. A., & Vallerand, R. J. (2003). On the hierarchical structure of self-determined motivation: A test of top-down, bottom-up, reciprocal, and horizontal effects. *Personality and Social Psychology Bulletin, 29*(8), 992–1004. https://doi.org/10.1177/014616720325329

Hajcak, G. (2012). What we've learned from mistakes: Insights from error-related brain activity. *Current Directions in Psychological Science, 21*(2), 101–106. https://doi.org/10.1177/0963721412436809

Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity. *Psychological Science, 19*(2), 103–108. https://doi.org/10.1111/j.1467-9280.2008.02053.x

Hajcak, G., Klawohn, J., & Meyer, A. (2019). The utility of event-related potentials in clinical psychology. *Annual Review of Clinical Psychology*, *15*(1), 71–95. https://doi.org/10.1146/annurev-clinpsy-050718-095457

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–197. https://doi.org/10.1016/j.bandc.2003.11.001

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*(2), 151–160. https://doi.org/10.1111/j.1469-8986.2005.00270.x

Härpfer, K., Carsten, H. P., Kausche, F. M., & Riesel, A. (2025). Enhanced Performance Monitoring as a Transdiagnostic Risk Marker of the Anxiety and Obsessive–Compulsive Spectrum: The Role of Disorder Category, Clinical Status, Family Risk, and Anxiety Dimensions. *Depression and Anxiety*, *2025*(1), 9505414. https://doi.org/10.1155/da/9505414

Holmes, A. J., & Pizzagalli, D. A. (2010). Effects of task-relevant incentives on the electrophysiological correlates of error processing in major depressive disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 119–128. https://doi.org/10.3758/CABN.10.1.119

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Iannaccone, R., Hauser, T. U., Staempfli, P., Walitza, S., Brandeis, D., & Brem, S. (2015). Conflict monitoring and error processing: New insights from simultaneous EEG–fMRI. *NeuroImage*, *105*, 395–407. https://doi.org/10.1016/j.neuroimage.2014.10.028

Jackson, F., Nelson, B. D., & Proudfit, G. H. (2015). In an uncertain world, errors are more aversive: Evidence from the error-related negativity. *Emotion (Washington, D.C.)*, *15*(1), 12–16. https://doi.org/10.1037/emo0000020

Khazanov, G. K., Ruscio, A. M., & Forbes, C. N. (2020). The positive valence systems scale: development and validation. *Assessment, 27*(5), 1045–1069. https://doi.org/10.1177/1073191119869836

Kim, E. Y., Iwaki, N., Uno, H., & Fujita, T. (2005). Error-related negativity in children: Effect of an observer. *Developmental Neuropsychology*, *28*(3), 871–883. https://doi.org/10.1207/s15326942dn2803_7

Kirschner, H., & Ullsperger, M. (2024). The medial frontal cortex, performance monitoring, cognitive control, and decision making. In Reference Module in Neuroscience and Biobehavioral Psychology. Elsevier. https://doi.org/10.1016/B978-0-12-820480-1.00048-6

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience, 21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: a theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology, 93*(3), 283–297. https://doi.org/10.1016/j.ijpsycho.2014.06.007

Legault, L., & Inzlicht, M. (2013). Self-determination, self-regulation, and the brain: Autonomy improves performance by enhancing neuroaffective responsiveness to self-regulation failure. *Journal of Personality and Social Psychology, 105*(1), 123–138. https://doi.org/10.1037/a0030426

Leue, A., Rodilla, C. C., & Beauducel, A. (2017). Worry-inducing stimuli in an aversive Go/NoGo task enhance reactive control in individuals with lower trait-anxiety. *Biological Psychology, 125*, 1–11. https://doi.org/10.1016/j.biopsycho.2017.02.003

Lewis, M., Mathur, M., VanderWeele, T., & Frank, M. C. (2020). The puzzling relationship between multi-lab replications and meta-analyses of the rest of the literature. *PsyArXiv*. https://doi.org/10.31234/osf.io/pbrdk

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*(1/2), 187-192. https://doi.org/10.2307/2333251

Lüdecke, D. (2019). esc: Effect size computation for meta analysis (Version 0.5.1). https://CRAN.R-project.org/package=esc

Luijten, M., Machielsen, M. W., Veltman, D. J., Hester, R., de Haan, L., & Franken, I. H. (2014). Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people with substance dependence and behavioural addictions. *Journal of Psychiatry and Neuroscience, 39*(3), 149–169. https://doi.org/10.1503/jpn.130052

Lutz, M. C., Kok, R., Verveer, I., Malbec, M., Koot, S., van Lier, P. A., & Franken, I. H. (2021). Diminished error-related negativity and error positivity in children and adults with externalizing problems and disorders: a meta-analysis on error processing. *Journal of Psychiatry and Neuroscience, 46*(6), E615–E627. https://doi.org/10.1503/jpn.200031

Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology: General, 129*, 43–60. https://doi.org/10.1037/0096-3445.129.1.43

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science, 14*, 47–53. https://doi.org/10.1111/1467-9280.01417

Maier, M., Bartoš, F., & Wagenmakers, E. J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods, 28*(1), 107–122. https://doi.org/10.1037/met0000405

Manoach, D. S., & Agam, Y. (2013). Neural markers of errors as endophenotypes in neuropsychiatric disorders. *Frontiers in Human Neuroscience, 7*, 350. https://doi.org/10.3389/fnhum.2013.00350

Maruo, Y., Schacht, A., Sommer, W., & Masaki, H. (2016). Impacts of motivational valence on the error-related negativity elicited by full and partial errors. *Biological Psychology, 114*, 108–116. https://doi.org/10.1016/j.biopsycho.2015.12.004

Meyer, A. (2017). A biomarker of anxiety in children and adolescents: A review focusing on the error-related negativity (ERN) and anxiety across development. *Developmental Cognitive Neuroscience, 27*, 58–68. https://doi.org/10.1016/j.dcn.2017.08.001

Meyer, A., Carlton, C., Chong, L. J., & Wissemann, K. (2019). The presence of a controlling parent is related to an increase in the error-related negativity in 5–7 year-old children. *Journal of Abnormal Child Psychology, 47*(6), 935–945. https://doi.org/10.1007/s10802-018-0503-x

Meyer, A., & Gawlowska, M. (2017). Evidence for specificity of the impact of punishment on error-related brain activity in high versus low trait anxious individuals. *International Journal of Psychophysiology, 120*, 157–163. https://doi.org/10.1016/j.ijpsycho.2017.08.001

Meyer, A., Lerner, M. D., De Los Reyes, A., Laird, R. D., & Hajcak, G. (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. *Psychophysiology, 54*(1), 114–122. https://doi.org/10.1111/psyp.12664

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28*(6), 487–495. https://doi.org/10.1016/0005-7967(90)90135-6

Moran, T. P., Schroder, H. S., Kneip, C., & Moser, J. S. (2017). Meta-analysis and psychophysiology: A tutorial using depression and action-monitoring event-related potentials. *International Journal of Psychophysiology, 111*, 17–32. https://doi.org/10.1016/j.ijpsycho.2016.07.001

Moser, J. S., Moran, T. P., Schroder, H. S., Donnellan, M. B., & Yeung, N. (2013). On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. *Frontiers in Human Neuroscience, 7*, 466. https://doi.org/10.3389/fnhum.2013.00466

Niu, Y., Li, Z., Pettit, J. W., Buzzell, G. A., & Zhao, J. (2023). Context and domain matter: the error-related negativity in peer presence predicts fear of negative evaluation, not global social anxiety, in adolescents. *Psychological Medicine, 53*(14), 6899–6909. https://doi.org/10.1017/S0033291723000466

Nuñez-Estupiñan, X., Berticelli, L. Z., de Almeida, R. M. M., & Gauer, G. (2022). Aversiveness of errors and the error-related negativity (ERN): A systematic review on the affective states' manipulations findings. *Cognitive, Affective, & Behavioral Neuroscience, 22*(4), 754–776. https://doi.org/10.3758/s13415-022-01002-2

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pailing, P. E., & Segalowitz, S. J. (2004). The error-related negativity as a state and trait measure: Motivation, personality, and ERPs in response to errors. *Psychophysiology, 41*(1), 84–95. https://doi.org/10.1111/1469-8986.00124

Palmwood, E. N., & Simons, R. F. (2021). Unsupportive romantic partner behaviors increase neural reactivity to mistakes. *International Journal of Psychophysiology, 170*, 12–19. https://doi.org/10.1016/j.ijpsycho.2021.09.008

Pasion, R., & Barbosa, F. (2019). ERN as a transdiagnostic marker of the internalizing-externalizing spectrum: A dissociable meta-analytic effect. *Neuroscience & Biobehavioral Reviews, 103*, 133–149. https://doi.org/10.1016/j.neubiorev.2019.06.013

Paul, K., Angus, D. J., Bublatzky, F., Wüllhorst, R., Endrass, T., Greenwood, L. M., ... & Pourtois, G. (2025). Revisiting the electrophysiological correlates of valence and expectancy in reward processing – A Multi-lab replication. *Cortex, 184*, 150–171. https://doi.org/10.1016/j.cortex.2024.12.017

Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, S. Y., Beste, C., Bland, A. R., Bradford, D. E., Busch, N. A., Clayson, P. E., Cruse, D., Dreber, A., Dumas, G., Ehinger, B., He, X., Hinojosa, J. A., Huber-Huber, C., Jack, B. N., Johannesson, M., … Mushtaq, F. (2021). #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex, 144*, 213–229. https://doi.org/10.1016/j.cortex.2021.03.013

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends in Cognitive Sciences, 10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C. T., Themanson, J. R., & Hillman, C. H. (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. Psychophysiology, 47(4), 767–773. https://doi.org/10.1111/j.1469-8986.2010.00974.x

Potts, G. F. (2011). Impact of reward and punishment motivation on behavior monitoring as indexed by the error-related negativity. *International Journal of Psychophysiology, 81*(3), 324–331. https://doi.org/10.1016/j.ijpsycho.2011.07.020

Proudfit, G. H., Inzlicht, M., & Mennin, D. S. (2013). Anxiety and error monitoring: the importance of motivation and emotion. *Frontiers in Human Neuroscience, 7*, 636. https://doi.org/10.3389/fnhum.2013.00636

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Riesel, A. (2019). The erring brain: Error-related negativity as an endophenotype for OCD—A review and meta-analysis. *Psychophysiology, 56*(4), e13348. https://doi.org/10.1111/psyp.13348

Riesel, A., Kathmann, N., Wüllhorst, V., Banica, I., & Weinberg, A. (2019a). Punishment has a persistent effect on error-related brain activity in highly anxious individuals twenty-four hours after conditioning. *International Journal of Psychophysiology, 146*, 63–72. https://doi.org/10.1016/j.ijpsycho.2019.09.014

Riesel, A., Klawohn, J., Grützmann, R., Kaufmann, C., Heinzel, S., Bey, K., ... & Kathmann, N. (2019b). Error-related brain activity as a transdiagnostic endophenotype for obsessive-compulsive disorder, anxiety and substance use disorder. *Psychological Medicine, 49*(7), 1207-1217. https://doi.org/10.1017/S0033291719000199

Riesel, A., Weinberg, A., Endrass, T., Kathmann, N., & Hajcak, G. (2012). Punishment has a lasting impact on error-related brain activity. *Psychophysiology, 49*(2), 239–247. https://doi.org/10.1111/j.1469-8986.2011.01298.x

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ, 340,* c221. https://doi.org/10.1136/bmj.c221

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese, 199*(3), 10969–11000. https://doi.org/10.1007/s11229-021-03276-4

Saunders, B., & Inzlicht, M. (2020). Assessing and adjusting for publication bias in the relationship between anxiety and the error-related negativity. *International Journal of Psychophysiology, 155*, 87–98. https://doi.org/10.1016/j.ijpsycho.2020.05.008

Saunders, B., Milyavskaya, M., & Inzlicht, M. (2015). What does cognitive control feel like? Effective and ineffective cognitive control is associated with divergent phenomenology. *Psychophysiology, 52*(9), 1205–1217. https://doi.org/10.1111/psyp.12454

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology, 10*, 813. https://doi.org/10.3389/fpsyg.2019.00813

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., McDonough, M., & Gillan, C. M. (2020). A dimensional investigation of error-related negativity (ERN) and self-reported psychiatric symptoms. *International Journal of Psychophysiology, 158*, 340–348. https://doi.org/10.1016/j.ijpsycho.2020.09.019

Shiels, K., & Hawk, L. W. (2010). Self-regulation in ADHD: The role of error processing. *Clinical Psychology Review, 30*(8), 951–961. https://doi.org/10.1016/j.cpr.2010.06.010

Spielberger, C., Gorsuch, R. L., & Lushene, R. E. (1970). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Stürmer, B., Nigbur, R., Schacht, A., & Sommer, W. (2011). Reward and punishment effects on error processing and conflict control. *Frontiers in Psychology, 2*, 335. https://doi.org/10.3389/fpsyg.2011.00335

Torrubia, R., Avila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences, 31*(6), 837–862. https://doi.org/10.1016/S0191-8869(00)00183-5

Unger, K., Kray, J., & Mecklinger, A. (2012). Worse than feared? Failure induction modulates the electrophysiological signature of error monitoring during subsequent learning. *Cognitive, Affective, & Behavioral Neuroscience, 12*(1), 34–51. https://doi.org/10.3758/s13415-011-0061-y

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., & Hasbroucq, T. (2003). Error negativity on correct trials: a reexamination of available data. *Biological Psychology, 64*(3), 265–282. https://doi.org/10.1016/S0301-0511(03)00097-8

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity' specific to errors?. *Biological Psychology, 51*(2-3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weinberg, A., Dieterich, R., & Riesel, A. (2015). Error-related brain activity in the age of RDoC: A review of the literature. *International Journal of Psychophysiology, 98*(2), 276–299. https://doi.org/10.1016/j.ijpsycho.2015.02.029

Weinberg, A., Meyer, A., Hale-Rude, E., Perlman, G., Kotov, R., Klein, D. N., & Hajcak, G. (2016). Error-related negativity (ERN) and sustained threat: Conceptual framework and empirical evaluation in an adolescent sample. *Psychophysiology, 53*(3), 372–385. https://doi.org/10.1111/psyp.12538

Weinberg, A., Riesel, A., & Hajcak, G. (2012). Integrating multiple perspectives on error-related brain activity: The ERN as a neural indicator of trait defensive reactivity. *Motivation and Emotion, 36*(1), 84–100. https://doi.org/10.1007/s11031-011-9269-y

White, E. J., Grant, D. M., Taylor, D. L., Frosio, K. E., Mills, A. C., & Judah, M. R. (2018). Examination of evaluative threat in worry: Insights from the error-related negativity (ERN). *Psychiatry Research: Neuroimaging, 282*, 40–46. https://doi.org/10.1016/j.pscychresns.2018.10.006

Wiswede, D., Münte, T. F., & Rüsseler, J. (2009). Negative affect induced by derogatory verbal feedback modulates the neural signature of error detection. *Social Cognitive and Affective Neuroscience, 4*(3), 227–237. https://doi.org/10.1093/scan/nsp015

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931