**Revisiting Defensive Motivation and the Error-Related Negativity: A Multi-Site Replication Study**

Amy R. Bland, Gary Dicks

Department of Psychology, Manchester Metropolitan University

Angelica DeFalco

Department of Psychology, Colorado State University, USA

Daniel E. Bradford,

School of Psychological Science, Oregon State University, USA

Valerie Bertaina-Anglade, Philippe L’hostis, Arzhelenn Guillo,

Biotrial, Core Lab, Rennes, France

Gilles Pourtois, Nan Wang,

Department of Experimental Clinical & Health Psychology, Ghent University, Ghent, Belgium

Zhong-Xu Liu

University of Michigan-Dearborn, Dearborn, Michigan, USA

Anna Fischer, Francesco Grassi, Anne Schacht, Annika Ziereis

Department for Cognition, Emotion and Behavior, University of Goettingen, Germany

Greg Hajcak,

Department of Psychology and Biomedical Sciences, Florida State University, USA

Dan Foti,

Purdue University, USA

Faisal Mushtaq

School of Psychology, University of Leeds

Yuri G. Pavlov

University of Tuebingen, Tuebingen, Germany

Not sure where you want me, but here my info:

Ingmar E.J. de Vries

1. Donders Institute, Radboud University, The Netherlands
2. Centre for Mind/Brain Sciences (CIMeC), University of Trento, Italy

\*Correspondence to Amy R. Bland, Manchester Metropolitan University, a.bland@mmu.ac.uk

**Competing Interests:**

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

**Funding:**

#EEGManyLabs is supported by the DFG (PA 4005/1-1) and the UK Biotechnology and Biological Sciences Research Council (BB/X008428/1). The funders had no role in study design or the decision to submit the work for publication.

**Author contributions**:

Author contributions are coded according to the CRediT taxonomy (Allen et al., 2014).

Conceptualization: YGP, FM, DEB, ARB

Data curation:

Formal Analysis:

Funding acquisition: YGP

Investigation:DEB, AD, ARB

Methodology:

Project administration: YGP FM

Resources:

Software:

Supervision: AB

Validation: AD

Visualization:

Writing – original draft: ARB, DEB, AD

Writing – review & editing:

**Abstract**

A large number of EEG studies have identified a distinct event related potential (ERP) component during error processing known as the Error-Related Negativity (ERN). In an influential study, Hajcak and Foti (2008) explored the idea that errors could trigger defensive motivational reactions and that the ERN might forecast the intensity of defensive reactions following errors. Using a flanker task, thirty-one college-aged participants responded to the direction of a central arrow with acoustic startle probes administered pseudo-randomly throughout. Hajcak and Foti’s (2008) findings indicated the ERN is indicative of individual variations in aversive reactions to errors. This has influenced understanding of the ERN being more than a simple error detection mechanism and sheds light on how people differ in their emotional responses to mistakes. Importantly, identifying how individual differences in aversive reactions manifest in ERN patterns could contribute to diagnostic strategies and personalized treatment approaches. As part of the #EEGManyLabs project, an international network of laboratories will test the replicability of the results from this influential study. The data will later be combined to compute global effect sizes of the ERN, startle potentiation, and their interaction. Collectively, these replications will help solidify the results from this highly-cited study. Given that the ERN is an integral part of a broader neural system responding to potentially threatening stimuli, this replication will provide a more solid foundation for our understanding of error processing and its relationship to defensive reactivity.

**Introduction**

One of the most widely studied event-related potentials (ERP) is the Error-Related Negativity (ERN). The ERN is characterized by a large deflection in the ERP waveform that peaks within ~100 milliseconds of an erroneous response (Gehring et al., 2018). Studies utilizing source localization suggest that the ERN is generated in the medial frontal cortex, and more specifically the anterior cingulate cortex (ACC; Dehaene et al., 2003; Holroyd et al., 1998; Miltner et al., 1997). The ERN has been observed across different stimuli and response modalities and is thought to reflect the activity of a generic response monitoring system (Bernstein et al., 1995; Dehaene et al., 2003; Falkenstein et al., 2001), signaling a need to increase cognitive control and make behavioral adjustments (Botvinick et al., 2001). After 25 years of investigation, the ERN’s prominence is clear, with new theories about its functional significance continuing to emerge. More recently, it has been listed as a core physiological marker of *Cognitive Control, Sustained Threat,* and *Reward Learning* in the Research Domain Criteria (RDoC) that forms the basis of National Institute of Mental Health’s (NIMH) efforts to understand symptoms relevant to psychopathology.

The ERN component is largely regarded as a neural index of response monitoring processes (Bernstein et al., 1995; Holroyd et al., 2003; Luu et al., 2000). However, as affective and motivational variables influence the magnitude of ERN, the ERN may relate to emotional and motivational aspects of error detection. Indeed, individuals and groups characterized by excessive concern over errors – for whom errors may be both more salient and more aversive – show enhanced ERN amplitudes (Riesel et al., 2013). Specifically, the ERN has been implicated as a biomarker of obsessive compulsive disorder (OCD; Gehring et al., 2000), worry (Hajcak et al., 2003) and extreme negative affective experience (Hajcak et al., 2004; Luu et al., 2000). Importantly, the enhanced ERN often observed, is rarely reflected in performance differences, suggesting that the magnitude of the ERN varies as a function of motivational states independent of behavioral measures (Weinberg et al., 2012). Therefore, rather than indexing error or conflict detection per se, the ERN may be sensitive to dynamically established goal states, and reflect the negative affective response to errors (Hajcak et al., 2005).

In the momentum of considering the ERN beyond simple error detection, Hajcak and Foti (2008) were the first to demonstrate a direct link between errors and defensive motivation. Arguing that error detection should prompt defensive motivation and rapidly activate reflexes and dispositions toward action that protect the organism, they sought to determine whether defensive responding would similarly be increased following errors, compared with correct responses. The authors utilized startle potentiation, a robust psychophysiological measure of defensive reactivity. This refers to an increase in the intensity of the startle reflex in response to a sudden, intense stimulus (typically a loud, sudden noise) under aversive or threatening conditions relative to neutral conditions. Hajcak and Foti (2008) hypothesized that startle responses would be larger (i.e., potentiated) after errors than after correct responses, given the aversive nature of error-making. Furthermore, they hypothesized that if the ERN reflects a motivationally relevant correlate of error processing, larger ERNs should predict greater startle potentiation following errors. To test their hypotheses, Hajcak and Foti (2008) recorded ERNs using EEG and startle responses using electromyographic (EMG) activity in thirty-one undergraduate students (26 male, 5 female). Their experimental paradigm included both predictable and unpredictable acoustic startle probes whilst simultaneously measuring error-related brain activity to determine whether individuals with larger ERNs would demonstrate greater startle potentiation following errors. Consistent with their hypotheses, the ERN appeared to be indicative of individual variations in aversive reactions to errors. Subsequent work by these investigators, and others, has led to the proposal that the ERN could be a potential neural marker for defensive motivation, with variance in the ERN reflecting individual differences in the degree to which errors are evaluated as threatening (Weinberg et al., 2016). This perspective aligns with the idea that the ERN could be part of a broader neural system involved in responding to potentially threatening stimuli.

Since its publication, Hajcak and Foti (2008)’s article has been cited over 450 times according to Google Scholar. Given the influence this result has had on the scientific study of error processing, testing its replicability is especially important to validate further findings that may have clinical implications. Indeed, the reliability and replicability of research findings is a widespread concern in the scientific community. It has raised questions about the robustness of many published studies and the overall credibility of scientific research. To ameliorate the state of study replication, the #EEGManyLabs project is seeking to shed light on the robustness of results in highly cited and influential electroencephalographic (EEG) studies (Pavlov et al., 2021). Replication is essential for ensuring that these EEG findings are robust, credible, and contribute meaningfully to the advancement of our understanding of psychological phenomena, which is particularly prudent for clinically-relevant results. Hajcak and Foti (2008) recruited only 31 participants which was typical in the field at that time. Yet, despite the interest in the ERN and defensive motivation, there have only been two attempts to replicate the original procedure, measures, and analyses. One successful replication was conducted by the original authors who found a significant correlation between ERN magnitude and error-potentiated startle (Riesel et al., 2013). However this was only evident in a subgroup of 16 participants exhibiting higher ERNs The second replication was an unsuccessful attempt by Lewis and Pitts (<https://osf.io/82bwv>). Lewis and Pitts failed to find a significant correlation between ERN and startle potentiation (r=-.25, *p*=.11, one-tailed). However, it is important to note that the small sample size of 51 participants, though larger than the original study sample size, may not have been sufficiently powered to detect the effect.

In addition to concerns about statistical power, the generalizability of past studies have been questioned. There is also a consistent bias towards studies taking place in the United States which has been met with a recent push towards globalized efforts (Ledgerwood et al., 2022). Therefore it is important to build upon attempted replications of this influential study by utilizing an international network of laboratories. Both the original study and Lewis and Pitts’ replication effort was conducted in the USA, therefore diversifying this original study’s sample (i.e., applying a global sample) will increase the generalizability of its results and, ultimately, better test whether the ERN is a marker of defensive motivation in the broader population.

The focus on replicating highly cited ERP studies is motivated by the field’s reliance on these typically smaller sample sizes and the complex data analysis methods that allow for researcher inconsistencies (Pavlov et al., 2021). As part of the #EEGManyLabs project (Pavlov et al., 2021), the current Registered Report proposes a close replication of Hajcak and Foti (2008), involving the collection of data from laboratories across UK, USA, France and Belgium to test the robustness of the ERN, in tandem with startle potentiation, as a biomarker of defensive motivation. Having a large international sample will provide clearer evidence into how this study design and results replicate and generalize. The data will later be combined to compute global effect sizes of ERN, startle potentiation, and their interaction. Collectively, these replications will place our understanding of error processing on firmer grounds.

Our primary hypothesis (H.1) is that the magnitude of the ERN correlates with the degree of startle potentiation following errors: participants with larger ERNs demonstrate greater startle potentiation following errors. Additionally (H.2), we will test whether startle responses will be larger after errors than after correct responses.

**Methods**

This manuscript adheres to the recommended open science practices for psychophysiological research outlined by Garrett-Ruffin et al. (2021). All study materials, including comprehensive details about each site, including equipment and electrode specifications, site-specific language, and participant recruitment procedures, and code for stimulus presentation and data processing will be stored in a project repository on the Open Science Framework (https://osf.io/fbqtm/). All raw EEG data will be shared publicly in a suitable repository (<https://gin.g-node.org/>). The OSF repository functions as a central hub, connecting to raw EEG data as well as data processing and analysis code.

This replication is a global initiative involving replicating labs that encompass multiple independent study sites geographically dispersed. At the time of the initial Stage 1 report submission, there are four distinct replicating labs. Each study site will secure approval from their local institutional review board/ethics committee to conduct the study and share de-identified data.

***Participants***

Participants will be recruited predominantly from undergraduate courses or the local community and will be required to give written informed consent before participating. The inclusion criterion is that participants must be 18 years of age or older with no known hearing problems and typical-to-corrected vision needed in order to properly complete the behavioral task. Individuals recruited from undergraduate courses will receive compensation in the form of course credit for their time, while those from the community will be remunerated in accordance with local policies governing participant compensation.

**Power Analysis**

A power analysis was conducted using G\*Power based on an effect size estimate from the original study for the central finding (H.1). The statistics from the original study showed a correlation between the amplitude of the ERN and the degree to which errors potentiated the startle response as *r* = −.38, *p*<.05. A total of 303 participants will be needed to achieve 90% statistical power, assuming that the effect size is half of what was observed in the original study (i.e., -.19), in a one-sided test. For H.2, regarding the effect stating that startle magnitudes are larger following errors than following correct trials (original study results: *t*(30) = 2.51; estimated dz = 0.451), the power analysis indicated that a sample of 57 participants is required. The anticipated data collection involves each replicating lab gathering EEG data from a minimum of 50 participants or as many as necessary to collectively gather data from at least 303 participants overall.

***Experimental Paradigm***

The experimental paradigm will closely follow the methodology of Hajcak and Foti (2008) where they utilized an arrowhead version (using < and > symbols) of the flanker task previously used in (Hajcak et al., 2005). All participants will perform one practice block of 30 trials where the startle probe will be delivered on 10% of all trials in the practice block. As described in Hajcak and Foti (2008), the task was presented as follows: On each trial, five horizontally aligned arrowheads were presented, and participants had to respond to the direction of the central arrowhead by pressing the left or right mouse button. On compatible trials, all five arrowheads pointed in the same direction (either left or right), and on incompatible trials, the central arrowhead pointed in the direction opposite the direction of the flanking arrowheads. Compatible and incompatible trials were equally frequent, and all stimuli were presented for 200 ms with an intertrial interval that varied randomly from 500 to 1000 ms. Participants performed eight blocks of 30 trials. At the end of each block, participants received performance feedback designed to encourage fast and accurate responding. If performance was 75% correct or lower, the message “Please try to be more accurate” was displayed; performance above 90% correct was followed by “Please try to respond faster”; if performance was between these levels, the message “You're doing a great job” was displayed (Hajcak & Foti, 2008).

The experiment was originally programmed using Presentation (Neurobehavioral Systems, Inc., [www.neurobs.com](http://www.neurobs.com/)) which has been adapted for use with PsychoPy software v2023.2.3 (<https://www.psychopy.org/>) and available from the project repository on the Open Science Framework (https://osf.io/fbqtm/). Tasks will be translated into local languages (English and French).

**Startle probe**

Startle will be elicited with a 105-dB burst of white noise,50-ms in duration and near instantaneous rise time presented 300 ms after the response. Startle probe delivery will be performed through headphones worn by the participant during the task which will be connected to a stimulation computer. Loudness will be assessed using a sound meter. Startle probes will be presented on 50% of error trials, on 50% of correct trials that followed errors, and on a random 4% of other correct trials. This meant that some startle probes were predictable (after both errors and correct trials that followed errors), whereas others were not (randomly selected correct trials). Accordingly, there were three trial types: *error trials*, *predictable correct trials* (i.e., correct trials on which the startle probe was predictable), and *unpredictable correct trials* (i.e., correct trials on which the startle probe was not predictable).

At a maximum of 105dB, the intensity of this sound is safely below levels at which there might be any risk of pain or physical damage as established by OSHA (Occupational Safety and Health Administration) and NIOSH (National Institute for Occupational Safety and Health) guidelines. Specifically, risk associated with noise exposure is reduced in the current experiment by limiting noise intensity to 105dB, limiting total noise exposure time to no more than 5s, and using broad spectrum noise (i.e., white noise). The portion of the experiment in which white noise bursts will be delivered to participants will last approximately 10 minutes. During this period, participants will be exposed to no more than 100, 50-millisecond bursts of white noise for a total of 5 seconds of exposure. The OSHA recommended limit for noise exposure at 105 dB is no more than 1 hour/day (OSHA section 191095).

***Neurophysiological Recordings***

In Hajcak and Foti’s (2008) study, continuous electroencephalographic (EEG) and electromyographic (EMG) activity was recorded using an ActiveTwo head cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, The Netherlands). Recordings were taken from 64 scalp electrodes based on the 10-20 system, as well as from two electrodes placed on the left and right mastoids. The replicating labs will be using one of the following EEG systems: (1) ANT Neuro eego mylab with 64-channel Waveguard Touch caps (UK) 2) Brain Products actiCHamp Plus (USA), (3) Grael 4K-EEG Compumedics (France) and BioSemi Active 64-channel (Belgium). Using elastic caps, all labs will record data with equivalent Cz, Fz, and Pz channels, and mastoids (TP9, TP10) positioned according to the extended 10/20 EEG system (Chatrian et al., 1985).

**Table 1. *Overview of EEG set-up and recording details at each replicating lab***

| **Participating University** | **Amplifier System** | **Electrode/Cap Model, Number EEG + EOG** | **Sampling Rate** | **Reference, Ground** | **Online Filter** | **Screen Type, Size, Ratio, Refresh Rate** | **Stimulus Presentation, Language** | **Buttons for task** | **Sound delivery system** | **EMG electrodes for FPS** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Manchester Metropolitan University, UK** | ANT Neuro eego mylab | 64-channel Waveguard Touch caps + ref/gnd + EOG | 1000Hz | M1,M2 | Anti-aliasing hardware filter | LED/IPS, 24 in, 1920x1080, 60 Hz | Psychopy (2023.2.3), English | Left and right mouse buttons | Headphones (Audio Technica ATH-M30x) and sound meter | External bipolar EMG channel |
| **Oregon State University, USA** | Brain Products actiCHamp plus | 32 channel actiCAP slim/snap | 2500Hz | Left mastoid (online); linked mastoids (offline); dedicated ground | Brain Products anti-aliasing hardware filter | LED ASUS TUF 23.8” 1080P Monitor (VG249Q1R) | Presentation | Black Box Toolkit response pad | Headphones (Sennheister HD 280 Pro) and sound meter | EEG cap electrodes in bipolar montage |
| **CORE Lab, Rennes, France** | Grael, Compumedics | 32 channel quik-cap | 512Hz | M1,M2 | BP filter: 0.1-70 Hz | IIYAMA Prolite E2083 HSD LED,, 19 in, 1600x900 resolution | Eprime, French | Chronos response device, Psychology software tools | Headphones (Sennheiser HD 25-1 II) and chronos device | External bipolar EMG channel |
| **Ghent University, Belgium** | Active two Biosemi | 64-channel, electro-cap including CMS & DRL, six external electrodes | 512Hz | Online CMS & DRL; offline linked mastoids | LP filter: 102.4 Hz | 19-inch CRT monitor (1600 x 1200 resolution at 75Hz) | Presentation | Keyboard | Headphones and artificial ear (GRAS 43AG-S2, Ear and Cheek Simulator with Kemar Pinna) |  |
| **University of Michigan-Dearborn** | Active two Biosemi | 64-channel, electro-cap including CMS & DRL, six external electrodes | 1024 Hz | Online CMS & DRL; Offline linked mastoids | Anti-aliasing hardware and digital software filter | ASUS ROG Swift PG248Q 24-inch (1920x1080), 144Hz | Psychopy 2022.2.4 | Keyboard | Headphones (Sennheiser HD 25); Creative blaster AE-7 sound card | Biosemi EMG/ECG (external) active channels |
| **University of Goettingen, Germany** | Active two Biosemi | 64-channels, electrode-caps including CMS & DRL, six external electrodes (mastoids, vertical and horizontal EOG), two external electrodes (startle response) | 1024 Hz | Online CMS and DRL, offline linked mastoids | Upper hardware limit: ⅕th of sampling rate as bandwidth-filter (204.8 Hz) | Alienware 25 Gamingmonitor AW2523HF, 24.5", 1920 x 1080 resolution, 255 Hz | Psychopy | Left and right mouse buttons | Bayerdynamic DT 770 PRO headphones | Biosemi EMG external active flat electrodes |

**Startle response**

Hajcak and Foti’s (2008) measured the startle response using standard procedures for assessing defensive reactivity (Bradley et al., 2005; Grillon et al., 1994). The electrooculogram (EOG) generated from blinks and eye movements was recorded from four facial electrodes: two approximately 1 cm above and below the participant's right eye, one approximately 1 cm to the left of the left eye, and one approximately 1 cm to the right of the right eye. The startle response was measured with two electrodes placed approximately 12 mm apart under the participant's left eye on the obicularis muscle.

**Data Processing**

All EEG data will be analyzed in EEGLAB (Delorme & Makeig, 2004) and processed using the pipeline that follows the original study as closely as possible (see Hajcak and Foti, 2008). The final code necessary for data processing at each of the sites will be posted online at the submission of the Stage 2 report. The pipeline for the original study is described below:

For the startle data, EMG activity was band-pass filtered (28–512 Hz; 24 dB/octave roll-off), rectified, then low-pass filtered at 30 Hz (24 dB/octave) and baseline-corrected. Individual trials were examined and rejected if the startle reflex began less than 20 ms following probe onset.

EEG data were re-referenced to the numeric mean of the mastoids and band-pass filtered with cutoffs of 0.1 and 30 Hz. The EEG was segmented for each trial, beginning 200 ms before the response and continuing for 800 ms. The EEG was corrected for blinks and eye movements using the method developed by Gratton et al. (1983). Specific intervals for individual channels were rejected in each trial using a semiautomated procedure, with physiological artifacts identified by the following criteria: a voltage step of more than 50.0 μV between sample points, a voltage difference of more than 300.0 μV within a trial, and a maximum voltage difference of less than 0.50 μV within a 100-ms interval.

Startle response magnitudes and latencies were quantified in terms of the peak in the 20- to 120-ms window after the presentation of the startle probe. Following the original approach, to determine whether the magnitude of the ERN predicts the increase in startle following errors, we will define startle potentiation for each participant as the startle magnitude after errors minus the average of startle responses on predictable correct and unpredictable correct trials. The ERN will be defined as the average activity at electrode FCz (or Cz if FCz is not available, as the ERN can be captured at Cz as well (Sandre, Banica, Riesel, Flake, Klawohn and Weinberg, 2020)), in the 0- to 100-ms time window following response onset in error trials, after applying single-trial baseline normalization using the -100 to 0 ms window before response onset as baseline.

As well as adhering closely to the data preprocessing protocol outlined in Hajcak and Foti (2008), in addition, we will also implement a more recent pipeline optimized for measuring EMG startle (Bradford et al., 2014). This involves: (1) high-pass filtering (4th order 28 Hz Butterworth filter, zero phase shift), (2) epoching from 50 ms pre-probe to 250 ms post-probe onset, (3) rectifying and smoothing (2nd order 30 Hz Butterworth low-pass filter, zero phase shift) the data. We will reject trials with values greater than +/- 20 μV in the 50 ms pre-probe to 10 ms post-probe window as artifact (i.e., unstable baseline). We will reject trials with mean amplitude less than -10 μV in the 100-250 ms post-probe window as artifact (i.e., movement artifact and baseline over-correction). Next, we will visually inspect figures of epochs of all processed startle data (all trials vs. accepted trials vs. algorithm rejected trials) and manually reject trials with excessive deflection in the baseline or post-probe windows not detected by automatic artifact detection. We will exclude participants who have lost >30 % of trials due to artifact rejection.

As well as adhering closely to the data preprocessing protocol outlined in Hajcak and Foti (2008), in addition, we will also implement a more recent pipeline optimized for measuring ERN (Clayson et al., 2021). This involves: (1) re-referencing to the numeric mean of the mastoids, (2) implementing a bandpass filter ranging from 0.1 to 30Hz; (2) employing a notch filter set at 50/60Hz (depending on the lab) to eliminate any electrical noise; (3) performing spherical interpolation of channel activity that remains constant or significantly deviates from the activity of other channels, determined through visual inspection of data and plotting channel spectra maps; (4) removal of extremely noisy data segments (for improving ICA performance) as detected by EEGLAB’s clean\_artifacts.m function, which uses the artifact subspace reconstruction (ASR) algorithm; (5) cleaning the data of ocular, muscular, or ‘bad’ channel artifacts with Independent Component Analysis (using the function ‘runica’ implemented in EEGLAB) and using SASICA (Semi-Automated Selection of Independent Components of the electroencephalogram for Artifact correction) plugin in EEGLAB (Chaumon, Bishop, & Busch, 2015). The following options will be enabled in SASICA: ‘Autocorrelation’ to differentiate muscle components (components reflecting brain data are known to be strongly autocorrelated), ‘Focal components’ (to determine bad channels), ‘Signal to noise ratio’ (to reject components with a low signal to noise ratio), and ADJUST (for detection of eyeblinks, and vertical and horizontal eye movements). The final decision to reject or retain components will be based on SASICA’s recommendations, as well as visual inspection of each component’s topography and overall component data (using EEGLAB’s data scrolling feature); (6) rejecting epochs deviating more than 3.29 standard deviations (SD) (Tabachnick & Fidell, 2007) from trimmed normalized means with respect to joint probability, kurtosis or the spectrum. The same epoching and baseline correction measures will be applied here as in the original pipeline.

**Statistical analyses**

To test H.1, we will use Pearson's correlation between the ERN and startle response. To test H.2, we will run 2 x 2 ANOVA of startle magnitude with factors predictability (predictable, unpredictable) and trial type (i.e., correct or error). The effect of trial type will be used to make the conclusion of whether startle responses are larger after errors than after correct responses. The ERN and startle responses will be statistically evaluated with R. Greenhouse-Geisser correction will be applied to *p* values associated with multiple-*df*, repeated measures comparisons.

***Procedure***

The procedure will follow, as closely as possible, the process described in Hajcak and Foti (2008) with any deviations stated explicitly. Participants will be recruited via Institutional Review Board or Ethics Committee approved channels, using approved language. Upon their arrival to the lab, they will receive a verbal description of the experiment, have an opportunity to ask questions, and provide informed consent.

Participants will then be prepared for the EEG and EMG recordings in the participant room. First, the EEG cap will be positioned on the participant’s head, followed by EOG facial electrodes which will be placed on the orbicularis oculi and two electrodes which will be placed under the left eye to record eyeblink startle response. Once all sensors are correctly placed, participants will have a chance to familiarize themselves with the task and response buttons with a practice block of 30 trials, which will not be included in analysis. The training session will last approximately 3 minutes.

Participants will then complete the main task comprising 8 blocks of 30 trials (240 total trials) with feedback text written in between each block, using the mouse button to respond. If the participant’s accuracy is below 75% or above 90%, the text will encourage the participant to make more accurate or faster responses, respectively.

As part of the broader initiative on EEG replicability (#EEGManyLabs), in addition to replicating the study, labs will also collect resting state EEG data and certain personality measures (<https://osf.io/sp3ck/>, Pavlov et al., 2021). The analysis of both EEG and personality data is not within the scope of the current study; however, they will be consolidated across sites for a future replication project, the findings of which will be reported separately. To facilitate this, participating labs will gather 8 minutes of resting state EEG, and participants will complete three brief questionnaires (translated into the local language where applicable), including the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990), the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), and the State Trait Anxiety Inventory Trait Version (STAI-T; (Spielberger et al., 1970). Following the Flanker task, labs collecting additional data will administer the Edinburgh Handedness Inventory (EHI; Oldfield, 1971), the Behavioral Inhibition and Approach System Scales (BIS-BAS; (Carver & White, 1994), the Center for Epidemiologic Studies Depression Scale (Events, 1977), and the Short Version of the Big Five Inventory (Gerlitz & Schupp, 2005) questionnaires to participants.

Upon completion, participants will be debriefed, compensated and released. In all, the task (including electrode preparation) should last about one hour.

**Evaluation of the Replication of Effects**

Replication success was defined in the #EEGManyLabs protocol (Pavlov et al., 2021). For each hypothesis separately, we will, first, compute effect sizes (H.1: Pearson’s r Fisher’s z-transformed, H.2: Cohen’s dz) for each individual lab and then combine all datasets in a random-effects meta-analysis (with labs as a random effect) using the REML estimator for random-effects variance. The replication will be considered successful if a statistically significant meta-analytic estimate (p < .02) across replicating labs is observed and if the effect is in the expected direction. We will report distribution of the weighted effect sizes, their 95% confidence intervals, heterogeneity (τ2). The metafor package (Viechtbauer, 2010) for R will be used for the meta-analyses.

If either H.1 or H.2 results in statistically non-significant effects, we will estimate evidence for the null hypothesis using a random-effects Bayesian meta-analysis. For this, we will use JASP (Love et al., 2019) with default priors (Cauchy prior with a scale of 0.707).

**References**

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1–2), 29–37.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). “ Where did I go wrong?” A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1312.

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.

Bradford, D. E., DeFalco, A., Perkins, E. R., Carbajal, I., Kwasa, J., Goodman, F. R., Jackson, F., Richardson, L. N. S., Woodley, N., & Neuberger, L. (2022). Whose signals are being amplified? Toward a more equitable clinical psychophysiology. *Clinical Psychological Science*, 21677026221112116.

Bradley, M. M., Moulder, B., & Lang, P. J. (2005). When good things go bad: the reflex physiology of defense. *Psychological Science*, *16*(6), 468–473.

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.

Clayson, P. E., Baldwin, S. A., Rocha, H. A., & Larson, M. J. (2021). 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.

Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schürhoff, F., Recasens, C., Martinot, M. L. P., Leboyer, M., & Martinot, J.-L. (2003). Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulate. *Proceedings of the National Academy of Sciences*, *100*(23), 13722–13727.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145.

Falkenstein, M., Hielscher, H., Dziobek, I., Schwarzenau, P., Hoormann, J., Sundermann, B., & Hohnsbein, J. (2001). Action monitoring, error detection, and the basal ganglia: an ERP study. *Neuroreport*, *12*(1), 157–161.

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.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (2018). The Error-Related Negativity. *Perspectives on Psychological Science*, *13*(2), 200–204. https://doi.org/10.1177/1745691617715310

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science*, *11*(1), 1–6.

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

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Grillon, C., Ameli, R., Goddard, A., Woods, S. W., & Davis, M. (1994). Baseline and fear-potentiated startle in panic disorder patients. *Biological Psychiatry*, *35*(7), 431–439.

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., McDonald, N., & Simons, R. F. (2003). Anxiety and error-related brain activity. *Biological Psychology*, *64*(1–2), 77–90.

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–197.

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/https://doi.org/10.1111/j.1469-8986.2005.00270.x

Holroyd, C. B., Dien, J., & Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. *Neuroscience Letters*, *242*(2), 65–68.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport*, *14*(18), 2481–2484.

Ledgerwood, A., Hudson, S. T. J., Lewis Jr, N. A., Maddox, K. B., Pickett, C. L., Remedios, J. D., Cheryan, S., Diekman, A. B., Dutra, N. B., & Goh, J. X. (2022). The pandemic as a portal: Reimagining psychological science as truly open and inclusive. *Perspectives on Psychological Science*, *17*(4), 937–959.

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*(1), 43.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

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

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*, *93*(3), 377–385.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory*.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA* (Vol. 724). Thomson/Brooks/Cole Belmont, CA.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

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.

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/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*, 84–100.

**Acknowledgements**

#EEGManyLabs core team and leading lab would like to thank the original authors, Greg Hacjak and Dan Foti, for generously providing crucial documentation, programming scripts and method information. The #EEGManyLabs project is funded by a DFG grant (PA 4005/1-1) provided to YGP and a UK Research and Innovation Biotechnology and Biological Sciences Research Council award (BB/X008428/1) to ARB and FM.