This is a Stage 1 Registered Report submitted to Cortex

Does facial expression processing require attention? A multi-lab replication of Eimer et al. (2003)

Ziyi Wang[1,2], Alejandro J. Estudillo[1], Ala Yankouskaya[1], Nazanin S. Andalibi[3], Amy R. Bland[4], Chris R. H. Brown[1], Donovan G. Govan[3], Tijl Grootswagers[3], Jennifer Gutsell[5], José A. Hinojosa[6], Jing Ke[7], Yixuan Ku[7], Bryan Q. Z. Leong[1], Claudia Poch[8], Ellen Seiss[1,9], Efe Soyman[10], Zhixing Sun[5], Manuel Varlet[3], Faisal Mushtaq[11,12], Yuri G. Pavlov[13,14], Xun He[1,9]*

[1]School of Psychology, Bournemouth University, UK

[2]Department of Psychology, Durham University, UK

[3]The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Australia

[4]School of Psychology, Manchester Metropolitan University, UK

[5]Department of Psychology, Brandeis University, USA

[6]Department of Experimental Psychology, Cognitive Processes and Speech Therapy, Complutense University of Madrid, Spain

[7]Department of Psychology, Sun Yat-sen University, China

[8]Nebrija Research Center in Cognition (CINC), Nebrija University, Spain

[9]Multimodal Immersive NEuro-sensing (MINE) Research Cluster, Bournemouth University, UK

[10]Department of Psychology, Koç University, Türkiye

[11]School of Psychology, University of Leeds, UK

[12]NIHR Leeds Biomedical Research Centre, UK

[13]Department of Psychology, The University of Alabama at Birmingham, USA

[14]Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany

*Corresponding author: xhe@bournemouth.ac.uk

Abstract

How the brain processes facial expressions and whether this process requires focal attention have been central topics of neurophysiological research on face processing. In a seminal study, using scalp-recorded event-related potentials (ERPs), Eimer, Holmes, and McGlone (2003) found that emotional and neutral faces were differently processed in the brain as early as 160 ms after stimulus onset. Importantly, these ERP differences disappeared when attention was directed away from the face stimuli, suggesting the necessity of focal attention in facial expression processing. These results have influenced the revision of the Bruce and Young model of face processing, and greatly motivated further research in the field. Yet, subsequent research did not always agree with these findings, and the study was never directly replicated. Hence, this study has been selected by #EEGManyLabs, a global and multi-laboratory replication project, for a high-powered direct replication across multiple laboratories. We hypothesise the replication of the original positive findings, including the ERP effects differentiating emotional and neutral faces and their dependence on focal attention. To address the debate on facial expression processing and the role of N170, we will also take this opportunity to reassess key negative findings of the original study, namely no difference between emotional and neutral facial expressions in N170 and no ERP differences among the six facial expressions. Replication or revision of these findings will confirm the necessity of focal attention in emotional face processing and the postulated biomarkers of basic emotion processing. This study will provide a reproducible benchmark on face processing.

Keywords: facial expression, face processing, visual attention, replication, #EEGManyLabs

Introduction

The ability to process facial expressions rapidly and accurately is central to human social communication (Anderson & Phelps, 2001; Fox et al., 2000; Öhman et al., 2001). Faces convey critical information about another's emotional state and intentions, and the way the brain responds to these visual cues has been a topic of extensive research over the past few decades (Ekman, Friesen, & Ellsworth, 1972; Niedenthal & Ric, 2017). Furthermore, research has also been trying to answer whether focal attention is needed for facial expression processing, which represents the level of automaticity in perceiving emotion in faces (Pessoa et al., 2002; Vuilleumier et al., 2001). A seminal study in this field by Eimer, Holmes, and McGlone (2003) demonstrated that several scalp-recorded event-related potential (ERP) activities (as early as 160 ms after stimulus onset) distinguished between laterally presented neutral and emotional faces when visual attention was allocated to faces. They also showed that the effect disappeared when attention was diverted by a demanding perceptual task. The findings suggested the necessity of focal attention in processing of facial expressions (e.g., Pessoa et al., 2002) and challenged the notion that emotional face processing occurs pre-attentively (e.g., Vuilleumier et al., 2001). Since its publication, this study has been delivering a significant impact on the theoretical development of emotional face processing. As of February 2026, this article had been cited over 630 times, consistently throughout the years.

The Eimer et al. (2003) study employed a within-subject design in which participants completed two tasks (emotion and line tasks). In the emotion task, participants were asked to discriminate between neutral faces and emotional faces (laterally presented) showing one of the six basic emotions (disgusted, fearful, happy, sad, surprised, angry) in each block. In the line task, participants were asked to judge the length of line segments closer to the central fixation while viewing the same face displays. This perceptual task demanded visuospatial attention away from the face stimuli. Their results showed ERP differences between the neutral and emotional face conditions (termed "valence effect" in the original publication; we changed this term to "emotion effect" in the current study given that this effect could also relate to arousal; Russell,1980), and more importantly, the modulation of the emotion effect by the involvement of focal attention. With focal attention (emotion task), significant ERP differences were found between emotional and neutral faces (not differentiating among the six emotions) at four time windows over frontal, central, temporal, and occipital scalp regions (with the earliest window being 160-215 ms in the frontal area). However, no such difference was found when focal attention was directed away from faces. These results contradicted the basic emotion theories, which suggest that the basic emotions are rapidly detected by automatic and distinct neural mechanisms operating independently of the involvement of voluntary attention (Ekman, 1992; Ortony & Turner, 1990). Meanwhile, the findings were in agreement with the Bruce and Young model in that structural encoding of faces (represented by the N170 component of ERP; Bentin & Deouell, 2000) and the processing of emotional facial expressions are separate processes in the brain (Bruce & Young, 1986; Eimer, 2011).

These findings were confirmed by Holmes, Vuilleumier, and Eimer (2003). Participants viewed arrays containing two faces and two houses, with faces displaying either a fearful or neutral expression. Spatial attention was directed to the faces or houses. Results showed stronger frontal positivity (approximately 100 ms post-stimulus) for fearful faces

relative to neutral faces when attention was directed to faces. Just as Eimer et al. (2003) showed, this difference was absent when attention was paid to the houses.

When faces are presented at the location of the fixation point, however, early emotion-related processing appears to proceed automatically. Holmes, Kiss, and Eimer (2006) presented face stimuli at the fixation, and found that the initial stage of emotional face processing (160-220 ms) occurred even when spatial attention was directed to a line task (for a review, see Eimer & Holmes, 2007). Similarly, Schindler et al. (2020) also found that the emotion modulation effect in N170 (which is at the same time stage of Eimer et al.'s early frontal activity, but occurs in the occipital area instead) was task-independent, highlighting the automaticity of early facial expression perception (specifically at the time of structural encoding of faces). While centrally presented faces can elicit early, pre-attentive responses, later sustained processing, including early posterior negativity (EPN, which represents an early attention mechanism enhancing brain responses to emotional information; Schindler et al., 2019; Schupp et al., 2006) and late positive potential (LPP, which indexes emotional stimulus evaluation and attentional control; Schindler et al., 2019; Schupp et al., 2006), may still require focused attention. Interestingly, even when participants did not have awareness of facial expressions (subliminal presentation), an early difference between neutral and fearful faces was found in the occipito-temporal area (Pegna et al., 2008).

Schindler et al. (2021) then manipulated the perceptual load of the non-face task to assess the amount of residual attention needed for the emotional face processing, as high perceptual load tasks can effectively exhaust attentional resources available for other processes (Lavie, 2005). They found that the emotion effects in early ERP activities (P1 and N170) were not modulated by load, whereas the later effect in EPN was dependent on attention. Together, these findings suggest that the earliest stages of facial expression analysis operate without capacity limits, whereas subsequent processing remains attention bounded. This central-peripheral dissociation is especially noteworthy because Eimer et al. (2003) placed their faces in the parafovea (2.2° from fixation centre-to-centre, 3.4° to the far edges) – an eccentricity that retains considerably higher acuity than the periphery which is beyond 5° from fixation (Larson & Loschky, 2009; Rayner, 1998) yet is already subject to attentional gating. Eimer et al.'s (2003) results thus mark a boundary between foveal and peripheral emotion processing.

Building on this evidence, several theoretical accounts describe graded effects of attention on emotional face processing. Pessoa (2008) and Pessoa and Adolphs (2010) proposed multiple levels of automaticity: subcortical routes can trigger rapid affective responses even when attention is limited, whereas ERP components that originate from cortical structures (such as P1 and N170) are modulated by available attentional resources. Luo et al. (2010) further suggested that coarse detection of some negative emotions (especially fear) can occur automatically in early neural responses (e.g., P1), while precise categorisation requires attentional investment. A review by Schindler and Bublatzky (2020) concurs: early responses are typically attention-independent, whereas sustained engagement with emotional content, particularly for complex or subtle expressions, depends on focal attention. This body of work therefore converges on a dual-process view: rapid, automatic detection of emotional salience is followed by attention-dependent elaboration.

The present study is part of the #EEGManyLabs project, a global and large-scale multi-laboratory replication programme (Pavlov et al., 2021). The goal of #EEGManyLabs is to undertake direct replications of highly influential studies with predefined sample sizes and high statistical power (90% power for 50% of the original effect sizes). Eimer et al. (2003) was selected for replication by a community of EEG experts for its high scientific impact.

The work by Eimer et al. (2003) was among the first to demonstrate that attention modulates the neural processing of emotional faces, and it has shaped much of the subsequent empirical and theoretical landscape. To our knowledge, no direct replication research has been carried out for this study. A high-powered, pre-registered direct replication is therefore warranted. Confirming (or refuting) the attentional modulation originally reported would provide a more secure foundation for debates about the automaticity of emotional face processing. Accordingly, the current study reproduces the original experimental protocol to reassess with enhanced statistical precision the role of spatial attention in the early and late cortical dynamics of facial-expression perception. There are two main hypotheses according to Eimer et al. (2003):

> **Hypothesis 1 (H1)**: In the emotion discrimination task (attention directed to faces), emotional faces will elicit enhanced ERPs relative to neutral faces, including:
>
> > **H1.1**: an enhanced early frontal positivity (original time window 160-215 ms);
> >
> > **H1.2**: an enhanced positivity at frontal and central sites (original time window 220-315 ms);
> >
> > **H1.3**: an enhanced negativity at lateral temporal and occipital sites (original time window 220-315 ms);
> >
> > **H1.4**: an enhanced positivity over frontal, central and parietal regions (original time window 320-495 ms);
> >
> > **H1.5**: an enhanced positivity over frontal, central and parietal regions (original time window 500-700 ms).
>
> **H2**: In the line discrimination task (attention directed away from faces), no emotion effect as in H1 will be found in any of these five ERP activities (correspondingly **H2.1** to **H2.5**).

A noteworthy result of the original study was that the N170 component was unaffected by emotional content in either task. N170 is widely recognised as an index of the structural encoding of faces (Bentin & Deouell, 2000; Rossion & Jacques, 2011). By showing the absence of any emotion effect in N170, Eimer et al. (2003) supported the hypothesis that emotion processing occurs independently from structural encoding – a conclusion consistent with several studies reporting similar N170 amplitudes across facial expressions (Fisher et al., 2015; Hermann et al., 2002; Kiss & Eimer, 2008; Sprengelmeyer & Jentzsch, 2006). Many other studies, however, have shown that N170 does distinguish emotional from neutral faces and can even differentiate between specific emotions (Batty & Taylor, 2003; Blau et al., 2007; Carlson & Reinke, 2010; Righart & De Gelder, 2006). Fear and anger, in particular, often elicited larger N170 amplitudes than neutral faces (Batty & Taylor, 2003; Calvo et al.,

2014; Faja et al., 2016; Leppänen et al., 2007, 2008; Pegna et al., 2008; Rossignol et al., 2005; Rellecke et al., 2012, 2013), a pattern confirmed by meta-analysis (Hinojosa et al., 2015). Some studies have even found stronger N170 responses to happy expressions (Luo et al., 2010; Righart & De Gelder, 2008). Collectively, these findings indicate that N170 can be sensitive to emotional information (for reviews, see Hinojosa et al., 2015; Schindler & Bublatzky, 2020).

Re-examining Eimer et al. (2003), the frontal positivity between 160 and 215 ms may correspond to the vertex positive potential (VPP), an ERP component sharing a neural generator with N170 and likewise indexing early face perception (Joyce & Rossion, 2005). The presence of a emotion effect and its attentional modulation in this window raises the possibility that N170 could exhibit similar effects when statistical power is sufficient.

Whether and how N170 responds to emotional faces therefore remains an open question. In the present high-powered replication, we will reassess the differences in N170 responding to emotional vs. neutral faces. Accordingly, following the original findings of Eimer et al. (2003), we will test an additional hypothesis:

> **H3**: N170 will show no emotion effect (i.e., no differences between emotional and neutral faces), regardless of the attentional manipulation.

In their original study, Eimer et al. (2003) reported no differences in the ERPs among the six basic emotions. This finding suggested that all emotions are processed in essentially the same way. Subsequent research, however, has shown that different emotions can be processed differently. Negatively valenced and potentially threatening expressions (e.g., fear and anger) tend to be more salient and are processed more rapidly and automatically than other emotions, a pattern known as the negativity bias (Ito et al., 1998; Smith et al., 2003). Luo et al. (2010) confirmed this bias, demonstrating that negative expressions, particularly fear, elicit an enhanced P1 component even when attention is not focused on the faces. Threatening faces likewise produce attention-independent enhancements of the EPN relative to neutral or friendly faces (Aguado et al., 2019; Rellecke et al., 2011; Schupp et al., 2004, 2006; Weidner et al., 2022). Some of these studies employed emotion categorisation tasks similar to that in the Eimer et al. (2003) study (Ito et al., 1998; Smith et al., 2003). Meanwhile, others either demanded more than categorising emotional and neutral faces and requested participants to identify facial expressions (Aguado et al., 2019; Luo et al., 2010; Weidner et al., 2022), or reduced the attentional engagement by studying facial expression processing in passive viewing conditions (Rellecke et al., 2011; Schupp et al., 2004). Therefore, it is unlikely that the differences in the findings could be explained by task differences. These observations motivate us to explore whether ERPs would differ across the six emotions in the current replication. Following the results of Eimer et al. (2003), we will test the final hypothesis:

> **H4**: ERPs do not differ across the six basic emotions in any of the measurements stated in H1 or N170.

Method

This Stage 1 Registered Report follows current open-science recommendations for psychophysiological research (Garrett-Ruffin et al., 2021). All study materials and preliminary processing codes are available on the Open Science Framework (OSF; https://osf.io/4n2t8/), which will also contain full information of each site and full analysis scripts. The forthcoming raw anonymised EEG data will be openly accessible to the public on G-Node (https://gin.g-node.org/EEGManyLabs) following the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016; Pernet et al., 2019) standard. Each participating site will obtain local ethics approval for collecting, anonymising, and publicly sharing data.

Power analysis and sample size

The original study's key results were several significant Task × Emotion interactions ("Task × Valence interactions" in the original publication's wording) that demonstrated enhanced ERP responses to emotional faces relative to neutral faces when focal attention was allocated to the faces (emotion task), but not when attention was directed towards line segments (line task) (Eimer et al., 2003; sample size $N = 14$). This attention-dependent emotion effect was observed in five ERP activities over four temporal phases, including an early frontal positivity (160-215 ms, $F(1,13) = 5.2$, $p < .05$), a frontocentral positivity and an occipitotemporal negativity (both at 220-315 ms, Fs > 7.2, ps < .02), and positivities over frontal, central and parietal regions (320-495 and 500-700 ms, Fs > 10.0, ps < .01).

Following the #EEGManyLabs recommendations (Pavlov et al., 2021) and recognising that ERP studies often overestimate effect sizes due to small samples (Clayson et al., 2019), we calculated the required sample size of the replication. The calculation was based on the weakest original significant interaction found in 160-215 ms ($F = 5.2$, corresponding to $\eta^2_p = 0.2857$). We targeted 90% power to detect an effect half of the original effect (i.e., $\eta^2_p = 0.1429$) at an alpha level of .02. With these settings, an a priori power analysis using MorePower 6.0.4 (Campbell & Thompson, 2012) indicated a sample size of 82.

However, this sample size puts up a high demand on each lab's data collection work and is potentially impractical. Therefore, we have decided to reduce the required sample size from each replicating lab to 41 (half of the initial estimation; this reduced sample size is not related to the number or replicating labs). Following the meta-analytic theory (e.g., Borenstein et al., 2021), with a fixed total sample size, distributing participants across sites and then meta-analysing their effect sizes yields the same point estimate and nearly identical precision, provided the between-site heterogeneity is accounted for (see also Paul et al., 2025). As such, reducing the per-lab sample from 82 to 41 will not compromise the meta-analytic power of the study as long as the overall sample is large enough. Each replicating lab will collect behavioural and EEG data from 41 participants. If any participants are to be excluded from the analysis, more participants will be tested to replace the excluded participants.

Sensitivity analysis

Eimer et al. (2003) did not find any difference in the N170 component, neither did they observe any differences across the six basic emotions. The current replication plans to further explore these patterns by including negative hypotheses (H3 and H4) based on the original study. To check whether the replication study will have enough power to statistically

assess weak effects and be able to avoid mistakenly drawing negative conclusions, we ran sensitivity analyses using MorePower 6.0.4.

So far, the replication study has six participating labs, each has agreed to collect data from 41 participants. This gives a total sample size of 246. For the emotion effect in N170 and its interaction with task (H3), with 90% power and alpha level of .02, we will be able to detect the effect as small as $\eta^2_p = 0.0510$ (a small effect size). For the facial expression effect and its interaction with the other factors (H4), the smallest detectable effect size is $\eta^2_p = 0.0159$ (a small effect size). Therefore, the planned analysis is capable of reliably detecting small effect sizes and will only show null findings if the effects are below these levels.

Recruitment criteria and participant information

Volunteers are eligible for the study if they are 1) aged 18-60 years, 2) with normal or corrected-to-normal vision, 3) self-reporting to have no neurological condition, and 4) without any skin allergies. Each replicating lab will record participants' age, handedness (left, right, ambidextrous), total years of education (the total number of years in formal schooling starting from elementary school), highest level of education, and length of time spent in the country (all self-reported).

Given that face processing shows own-race bias, a perceptual and social bias leading to better responses towards own-race individuals (e.g., Anzures et al., 2013), we will also record each participant's ethnic origin or ancestry and race. This is important because all the face images in the study are of white individuals. These questions and answer options (Appendix 1) are adapted from the recommendation of Joint Commitment for Action on Inclusion and Diversity in Publishing for its well-known capability to capture inclusivity and diversity (Royal Society of Chemistry, 2020). The ethnicity/race information is not part of the planned analysis but will be kept in case of need to explore any potential contribution from ethnicity/race. For this reason, the option "prefer not to disclose" is removed from the list of answer options. Participants will have to withdraw from the study if they do not want to disclose information about their ethnic origin (ancestry) or race. Gender information will also be collected using this document, with the "prefer not to disclose" kept as an answer option.

Stimuli

In Eimer et al. (2003), 70 face images of ten individuals (six female, four male) from Pictures of Facial Affect (POFA; Ekman & Friesen, 1976) were used. Each image of any individual showed either the neutral expression or one of the six basic emotional expressions (angry, disgusted, fearful, happy, sad, surprised). All faces were in greyscale against a black background with external features removed. In the current replication, we will use the same 70 images. Provided by the lab of the original study, these are 8-bit greyscale bitmap images (100×142 pixels, portrait layout). To be authentic to the original stimuli, these images will be used in the replication despite the relatively low image resolution. Viewing distance in each replicating lab is determined to ensure the face images will extend the visual angles as reported in the original study. To avoid clearly noticeable quality degradation, we will use the images in their original display size if possible. If using the original image size requires a viewing distance less than 50 cm or is not allowed by the lab setup, we will stretch to enlarge

the images but only up to 150% of the original size. So far, all labs have confirmed 100% scaling (no stretching) as summarised in Table 1.

[insert Table 1 here]

The stimulus displays (Figure 1) comprise two identical face images and a pair of vertical line segments on a black background, all horizontally aligned with the position of a fixation cross. Each side of the screen contains one face and one line segment. Each face, either showing a neutral expression or one of the six basic emotions, subtends a visual angle of 3.4×2.4° (portrait layout) and is presented 2.2° to the left or right of the fixation (centre-to-centre). The line segments are 0.1° wide and positioned 0.4° to the left and right of the fixation, and can be both short (0.4°), both long (0.5°), or of different lengths (0.4 and 0.5°). The fixation cross remains at screen centre throughout the experiment.

[insert Figure 1 here]

Procedure

The experiment was programmed in PsychoPy v2023.1.2 (Peirce et al., 2019), following the information provided in Eimer et al. (2003) and further information provided by the corresponding author (Prof. Martin Eimer) of the original study. The full program, along with a README file, is openly available on the project's OSF repository (https://osf.io/4n2t8/). Because the replication will be conducted across replicating labs, detailed information about each lab's monitor (model, refresh rate, resolution), viewing distance, and face image scaling factor is summarised in Table 1. All replicating labs will make necessary changes to the program, including translating or rewording of the instructions for their participants, changing the method of manual responses, changing the trigger port address, changing the trigger values for block start and end, and slight tweaking for their own PsychoPy version. As stated above, each lab will try to stay with a 100% scaling factor if the hardware and lab setup allow. All replicating labs' experiment programs will also be available on the repository.

The experiment has 24 blocks, each of which has 80 trials. Participants will complete two discrimination tasks (emotion and line tasks), each comprising 12 consecutive blocks, with the task order counterbalanced across participants. Throughout the experiment, participants will be instructed to keep their eye gaze on the fixation cross and to minimise eye blinks and eye movements. In each block, no matter what the task is, neutral face pairs randomly appear on half of the trials, with emotional face pairs randomly appearing on the other half. Each of the six facial expressions (angry, disgusted, fearful, happy, sad, surprised; i.e. six block types) is presented on the emotional face trials for two blocks, with the order of block types randomised across participants. The length of each line segment in the display varies independently (short and long lines are equally likely to appear on either side of the display).

Each trial presents the stimulus display for 300 ms, followed by a 2,000-ms interval before the next trial starts (Figure 1). In the emotion task, participants will be asked to discriminate between emotional and neutral face pairs and ignore the line segments, and to respond as quickly and accurately as possible with left- or right-hand keypresses ('z' or 'm') for neutral or emotional faces. In the line task, participants will determine whether the line segments are identical or different in length while ignoring the faces, and to respond quickly

and accurately by pressing ‘z’ or ‘m’. The response-hand mapping is counterbalanced between tasks and across participants. After each block, feedback on participant performance (mean accuracy and RT) will be provided on the screen.

Before running the experimental blocks, each participant will go through a practice block for each task. Participants will stop practising and proceed to the experimental blocks if they have reached 80% accuracy in a practice block in each task, as in the original study (personal communication). If a participant does not reach the 80% accuracy level after one practice block, up to two more practice blocks will be carried out. If a participant’s mean accuracy in a practice block remains below 80% after three practice blocks for any task, this participant will not proceed to the experimental blocks.

EEG data acquisition

Replicating labs will record EEG data with the setups listed in Table 1 and following the EEG preparation procedure guidance video (available on https://osf.io/4n2t8/). Table 2 summarises the EEG data acquisition, pre-processing, and measurement pipelines. All data will be acquired with the specified filter and sampling rate settings. Each lab can have a different number of electrodes, but must be able to record from these key locations: frontal pole (Fpz; Fp1 instead if Fpz is not possible), frontal (F3, Fz, F4), central (C3, Cz, C4), parietal (P3, Pz, P4), lateral temporal (P7, P8), and lateral occipital (PO7, PO8), horizontal electrooculograph (HEOG; either as a single bipolar channel or as two monopolar channels), and two separate earlobes (if earlobes are not possible, separate mastoids or TP9 and TP10 instead). To maximise the accuracy of the independent component analysis (ICA), each lab is encouraged to record EEG from as many channels as possible (see the “advanced pipeline” subsection below). The collected data will be transferred to the leading lab for data pre-processing and statistical analysis, strictly following the pre-registered pipelines. The raw anonymised data will also be shared on G-Node (https://gin.g-node.org/EEGManyLabs) following the BIDS standard.

[insert Table 2 here]

In Table 2, next to the pipeline employed in the original study (Eimer et al., 2003), two pipelines are listed. First, there is the “replication pipeline” in which we stay as close as possible to the protocol of the original study. The other pipeline (“advanced pipeline”) is a revised pipeline informed by the current EEG research practice. The EEG data acquisition steps remain the same in the two pipelines.

Despite that Eimer et al. (2003) mentioned linked-earlobe reference for data acquisition, the original study in fact used the left earlobe as the online reference, and later digitally re-referenced the recordings to the averaged earlobes (personal communication). In EEG research, linked-earlobe referencing is not advisable due to the current shunting problem (e.g., Ollikainen et al., 2000). With physically linked electrodes, an electric current between the two electrodes can occur. This can distort the EEG recordings and reduce hemispheric differences in the results (Luck, 2014; Nunez & Srinivasan, 2006; Regan, 1989). Therefore, as in the original study, the replication will use a single-electrode online reference. As the physical reference is a built-in and non-customisable component in some EEG systems, each replicating lab can determine their own online reference, with the single-earlobe or single-mastoid option preferred if possible. To be consistent with the original

study, data from both earlobes or mastoids need to be obtained (as either an online reference or a monopolar channel), so that they can be used for offline re-referencing. If a replicating lab cannot apply earlobe or mastoid electrodes, TP9 and TP10 will be used in place of earlobes (Clayson et al., 2021; Jovanović et al., 2025).

Two settings are changed from the original study. First, Eimer et al. (2003) used an online bandpass filter of 0.1-40 Hz. To avoid unnecessarily removing data and to keep as much as possible in the replication study, no online filter will be applied unless a replicating lab's EEG system does not have the capability of DC recording. In this case, a 0.1 Hz highpass filter will be applied. The 0.1-40 Hz bandpass filtering will be achieved offline with a digital filter (i.e., 0.1-40 Hz bandpass filter if no online filter is used; 40 Hz lowpass filter if a 0.1 Hz online highpass filter is used). Second, a higher sampling rate (500 or 512 Hz, as seen fit for each individual EEG system) will be used (200 Hz in the original study) to provide a higher temporal resolution.

EEG data pre-processing and ERP measurement

All EEG data will be imported into EEGLAB (Delorme & Makeig, 2004) using data importation plugins and processed using two pipelines (Table 2), i.e., a direct replication pipeline that follows the original study as closely as possible, and an advanced pipeline following more recent standards. The pre-processing codes for both pipelines are available online (https://osf.io/4n2t8/). These codes have been tested with pilot data collected from several replicating labs for quality assurance and the choice of pre-processing parameters.

Replication pipeline

The data will be re-referenced to the average voltage of earlobes (or averaged mastoids, or averaged TP9 and TP10 if EEG data from earlobes are unavailable) before being down-sampled to 200 Hz (the same as in the original study). A bipolar HEOG channel will be generated if HEOG data were acquired as separate monopolar channels. A digital bandpass filter of 0.1-40 Hz will then be applied. The filtered data will be segmented into epochs ranging from 100 ms before to 700 ms after stimulus onset, then baseline corrected for the 100-ms pre-onset period. Afterwards, artefact rejection will be carried out to remove epochs contaminated by horizontal eye movements (exceeding ±30 μV at the HEOG channel), eye blinks (exceeding ±60 μV at Fpz or Fp1 if Fpz is unavailable), and other artefacts (exceeding ±80 μV at any other key electrodes, including F3, Fz, F4, C3, Cz, C4, P7, P3, Pz, P4, P8, PO7, and PO8). If more than 50% of epochs in any of the two tasks from a participant are to be removed by this step, this particular participant will be removed from any further analysis. We will apply the 50% threshold to each task separately, because the block design of the two tasks may lead to poor quality data in one task only, disproportionately affecting results from one task. Finally, accepted epochs carrying correct responses will be averaged to create ERP waveforms for each condition (i.e., each of the 24 task, block type, and emotion combinations).

For ERP quantification, we will use the same channels and time windows reported in Eimer et al. (2003) for all the ERP activities. Specifically, N170 will be quantified as mean amplitude at P7 and P8 (labelled T5 and T6 in the original study) over the time window of 160-200 ms. Five successive time windows will then be used to quantify mean ERP amplitudes over five scalp regions separately, namely frontal (F3, Fz, F4), central (C3, Cz,

C4), parietal (P3, Pz, P4), temporal (P7, P8), and lateral occipital (PO7 and PO8; labelled OL and OR in the original study) channels. These time windows are 120-155, 160-215, 220-315, 320-495, and 500-800 ms, which cover the periods showing emotional expression differences in studies predating the original study (Eimer & Holmes, 2002; Holmes et al., 2003).

Advanced pipeline

The data will be re-referenced and digitally filtered as above, but not be down-sampled. Afterwards, we will apply the ICA method (Onton et al., 2006). For this purpose, we will create a copy of the dataset and highpass filter it at 2 Hz (Klug & Gramann, 2021). Then, any data with no event triggers for longer than 5,000 ms will be removed. Independent component (IC) weights will then be computed with the Adaptive Mixture ICA (AMICA) method (Palmer et al., 2012), and be transferred to the original data. Another data copy will be created with the 2-Hz highpass filter for IC classification, which will employ ICLabel (Pion-Tonachini et al., 2019) to automatically identify IC sources. ICs identified as eye, muscle, heart, line noise, and channel noise with more than 80% probability will be flagged for rejection. These flagged ICs will be removed from the original dataset.

After the ICA procedure, the dataset will be segmented into 800-ms epochs (-100 to 700 ms relative to stimulus onset), baseline corrected, and artefact rejected. For artefact rejection, in addition to the three criteria above (eye movements, blinks, strong artefacts), trials showing prolonged weak activity (absolute voltage below 0.5 μV for more than 200 ms) at any of the key electrode locations (F3, Fz, F4, C3, Cz, C4, P7, P3, Pz, P4, P8, PO7, PO8) will also be rejected. As in the replication pipeline, a 50% epoch rejection threshold for each task will be used for participant removal. Average ERP waveforms will then be created based on epochs with correct behavioural responses.

ERP waveforms will be quantified at the same selection of channels as in Eimer et al. (2003), including P7/P8 for N170, and F3/Fz/F4, C3/Cz/C4, P3/Pz/P4, P7/P8, and PO7/PO8 for the other five ERP activities. To determine the analysis windows reliably and objectively, the collapsed-localiser method (Luck & Gaspelin, 2017) will be implemented separately for each replicating lab's dataset. From each lab's dataset, a localiser waveform will be generated for each electrode cluster (e.g., F3/Fz/F4) by averaging over these electrodes in the group, and across all conditions and participants. Then, for each localiser waveform, the local peak amplitude will be determined within the corresponding analysis window in the original study. For instance, the N170 peak amplitude will be measured between 160 and 200 ms in the localiser waveform averaged over P7/P8 across all conditions and participants. Following this protocol, the other five peak-searching windows will be 120-155, 160-215, 220-315, 320-495, and 500-800 ms (identical to the analysis windows in Eimer et al., 2003), each of which will be applied to the five electrode clusters separately (F3/Fz/F4, C3/Cz/C4, P3/Pz/P4, P7/P8, and PO7/PO8).

For each ERP activity, once the peak amplitude has been determined, the activity boundaries will be set at 50% of the peak amplitude. The time window for ERP quantification will be the period between the two boundary time points.

Planned statistical analyses

We will first check the overall accuracy and miss rate of all participants. Following the criterion used in the practice blocks, any participants showing an accuracy level lower than 80% in any of the two tasks will be excluded from further analysis. Furthermore, as each trial requires a response, if a participant fails to respond to more than 10% of all trials in each task, this participant will also be removed from further analysis.

Given the within-subjects design, the behavioural and ERP data will be statistically assessed with repeated-measures analyses of variance (ANOVAs; with Holm correction applied to all post hoc comparisons) and paired *t*-tests. The Greenhouse-Geisser correction will be applied to the ANOVAs whenever the sphericity assumption is violated. Outputs with *p* values smaller than .02 will be accepted as rejection of the null hypotheses. All analyses will be complemented by their corresponding Bayesian equivalent with a Cauchy prior width of 0.707 for fixed effects (except otherwise specified for H3). The Cauchy prior distribution (a heavy-tailed probability distribution) is often used as the default prior distribution for Bayesian analyses as it is suggested to be the simplest option satisfying consistency requirements for hypothesis testing (Jeffreys, 1961; Wetzels et al., 2012). The Bayesian factors will be quantified by comparing across matched models. We will descriptively interpret the Bayesian factor results using the categories summarised by Lee and Wagenmakers (2014). For the BF_{10} value evidencing the alternative hypothesis (the opposite of the null hypothesis), 1 means no evidence in favour of the alternative hypothesis; 1 to 3 is anecdotal evidence; 3 to 10 is moderate evidence; 10 to 30 is strong evidence; 30 to 100 is very strong evidence; above 100 is extreme evidence. The reciprocal of BF_{10} (i.e., BF_{01}) will suggest the corresponding evidence in favour of the null hypothesis.

If the frequentist and Bayesian analysis results diverge, we will favour the frequentist interpretation first as it is a direct replication of the original study. The Bayesian results will complement the frequentist interpretation.

Behavioural analysis

Differences in reaction time (RT) and accuracy between the emotion and line tasks will be first assessed using two paired *t*-tests. For each task, the RT and accuracy data will then be analysed using ANOVAs and Bayesian ANOVAs as stated above. In the emotion task, a two-way ANOVA will be conducted with within-subject factors Emotion (emotional vs. neutral) and Block Type (the six facial expressions). For the line task, an additional factor (Line Type with two levels identical lines vs. different lines) will be included, resulting in a three-way ANOVA.

ERP analysis

For each ERP activity resulting from each of the two pre-processing pipelines, the analysis of mean ERP amplitudes will start with a three-way ANOVA. The three factors are Task (emotion vs. line), Emotion (emotional vs. neutral), and Block Type (the six basic facial expressions). Following any significant main effects of Block Type in these ANOVAs, post hoc tests using Holm corrections will be used. Consistent with the original study, planned two-way ANOVAs will be performed for each task (irrespective of the Task x Emotion interaction outcome). Corresponding Bayesian ANOVAs will also be implemented using a Cauchy prior width of 0.707 except for H3. For H1 and H2 (predictions of positive findings), frequentist analysis outputs with $p < .02$ in the same direction as in the original study will be

accepted as successful replications. $BF_{10} \geq 3$ in the same direction as in the original study and $BF_{10} \leq 0.333$ will be accepted as evidence for the alternative and null hypothesis, respectively. For H3 and H4 (predictions of negative findings), results showing $p < .02$ or $BF_{10} \geq 3$ in either direction and $BF_{10} \leq 0.333$ will be accepted as evidence for the alternative and null hypothesis, respectively.

Results confirming H1 and H2 should show a significant Task x Emotion interaction in the three-way ANOVA. Planned two-way ANOVAs should demonstrate significantly higher amplitudes for the emotional than neutral faces in the emotion task (Emotion main effect) in the five ERP activities, including frontal positivity 160-215 ms (H1.1), frontocentral positivity 220-315 ms (H1.2), occipitotemporal negativity 220-315 ms (H1.3), frontal-central-parietal positivity 320-495 ms (H1.4), and frontal-central-parietal positivity 500-700 ms (H1.5). The corresponding Emotion main effects are expected to turn out in favour of the null hypothesis in the line task (H2.1 to H2.5).

For H3, no significant Task x Emotion interaction is expected in the frequentist three-way ANOVA. The following two-way ANOVAs should also show no main effect of Emotion in either the emotion or line task. As in the original study, frequentist paired *t*-tests will be performed to compare the emotional and neutral faces in the emotion task, separately for the six basic emotions (i.e., the six Block Type levels). No significant differences are expected to be found in these *t*-tests.

In the Bayesian ANOVAs for H3, we will run the same three-way ANOVA, followed by two-way ANOVAs for the two tasks separately. One key step we will take is to set the effect size prior following the study of Hinojosa et al. (2015). Their meta-analysis found a significantly stronger N170 in response to emotional faces than to neutral faces, with a mean Cohen's d of 0.3325. Hence, we will set the effect size prior of the Emotion main effect in the Bayesian ANOVA to this level (Cauchy prior width r = 0.3325). For consistency, the same prior will be used in the follow-up two-way ANOVAs for the two tasks separately, despite that Hinojosa et al. (2015) reported significantly larger effect size in the non-direct task (attention directed away from faces, d = 0.3574) than in the direct task (attention directed to faces, d = 0.2025). In this process, we will keep the priors for all other fixed effects to 0.707. Evidence for the null hypothesis will confirm H3.

Furthermore, Hinojosa et al. (2015) found different effect sizes between emotional and neutral faces across various emotions, namely angry vs. neutral (Cohen's d = 0.2062), fearful vs. neutral (0.4219), and happy vs. neutral (0.2552). The effects between disgusted and neutral (0.4217, with a large confidence interval that included zero) and sad and neutral (0.0769) were not significant in their meta-analysis. To check the robustness of the results, we will repeat the Bayesian ANOVA twice with different Emotion effect priors - the smallest (0.2062) and the largest (0.4219) significant effect size reported by Hinojosa et al. (2015), respectively targeting very small and larger effect sizes in the literature. This will help us more confidently reveal whether there is an emotion effect in N170 and the plausible effect size. In total, there will be six two-way Bayesian ANOVAs (three for each task). In the emotion task, as in the plan for the frequentist analysis, we will follow up each ANOVA with six Bayesian paired *t*-tests for the six emotions, all using the same prior setting as in the preceding ANOVA.

H4 is relevant to all the analyses planned above. The absence of the Block Type difference (i.e., no differences among the six emotions) in all the frequentist and Bayesian ANOVAs described above can be considered support for H4. However, it is important to note that, in the design of the original study, each block contains both emotional and neutral faces with equal probabilities. The inclusion of the neutral face data (50% of the dataset) in the analysis will reduce the power to detect facial expression differences. Hence, to directly test H4, instead of relying on the exact replication analyses mentioned above, we will run dedicated ANOVAs (both frequentists and Bayesian) with only the emotional face data to directly compare the six emotions. This will remove the Emotion (emotional vs. neutral) factor and turn the analysis into a two-way ANOVA of Task x Emotion Type. The factor Emotion Type has six levels corresponding to the six emotion types (we will not use the name "Block Type" anymore given that no full data from any block will be used).

Furthermore, it is difficult to directly run a meta-analysis on the effect of Emotion, which has six levels and does not have direction information. Therefore, we will adopt a different approach for the assessment of H4. For each of the six ERP activities (the five listed in H1 plus N170), we will run the frequentist Task x Emotion Type ANOVA (still only using the emotional face data) on the whole sample across all replicating labs, then follow up with a planned one-way ANOVA (only frequentist) for each task. This is to check the possibility of any evidence supporting the cross-emotion differences in ERPs. Any significant ($p < .02$) one-way ANOVAs will be accepted as indications of potential ERP differences among the emotions. We will then analyse individual replicating labs' data (emotional trials only) with paired *t*-tests (both frequentist and Bayesian) contrasting all emotion pair combinations. The Bayesian prior will be set to 0.707.

Meta-analysis

To evaluate the replication success of the original study, meta-analyses of ERP data will be carried out following the #EEGManyLabs protocol (Pavlov et al., 2021). Specifically, we will run meta-analyses for the ERP data directly corresponding to the hypotheses, and will not run meta-analyses for behaviour data or any other ERP data. Table 3 summarises the individual analysis results that will be entered into the meta-analyses. The frequentist analysis results will be entered into frequentist meta-analyses, whereas the Bayesian analysis results will be put into Bayesian meta-analyses. In cases where the frequentist and Bayesian analyses diverge, we will prioritise the frequentist interpretation complemented by the Bayesian results as evidence for or against the effect.

[insert Table 3 here]

From the frequentist analysis results in each replicating lab's dataset, effect size (Cohen's d) and its standard error (SE) will be calculated for each hypothesis and put into a random-effects meta-analysis using the REML method, with the random effect being the replicating labs. This approach will address the heterogeneity in the datasets (due to differences in EEG systems, research practice, ethnicity/race, cultural influence, language being used, and samples). For each positive hypothesis, the replication will be accepted as successful if the meta-analytic estimate across all replicating labs is significant ($p < .02$) and in the hypothesised effect direction. For each negative hypothesis, we define successful

replication as a non-significant meta-analytic estimate ($p > .02$). The values to be reported include point estimate, distribution of the weighted effect sizes, 98% confidence intervals, and heterogeneity (τ^2).

For the Bayesian meta-analyses of replication results, we will extract standardised effect sizes (Cohen's d) and their associated uncertainty (credible intervals) from each individual dataset. These will be entered into a Bayesian random-effects meta-analysis to estimate the pooled effect size and between-study heterogeneity. For each meta-analysis, we will specify a normal prior centered at 0 with SD = 0.5 on the overall effect and a weakly informative prior on the heterogeneity parameter (half-Cauchy with a scale of 0.5). Replication success for positive findings 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. Replication success for negative findings (null hypotheses) will be demonstrated by $BF_{10} \leq 0.333$. This approach will allow us to quantify the cumulative evidence across replication datasets while accounting for uncertainty. We will report the pooled effect (μ), posterior distribution, 98% confidence intervals, and heterogeneity (τ^2).

Acknowledgements

We would like to thank Dr Rebecca Nako and Prof Martin Eimer for providing the stimuli and presentation protocol of the original study. We also thank Prof Martin Eimer for commenting on the manuscript. Our thanks also go to Beyza Tunç, Samantha Maramo, and Niamh Barry (Bournemouth University) for their help with pilot data collection and the video recording of the data acquisition protocol.

Conflict of Interest

The authors declare no conflict of interest.

Funding

The #EEGManyLabs project is supported by a Deutsche Forschungsgemeinschaft (DFG) grant to YGP (PA 4005/1-1) and a UK Research and Innovation Biotechnology and Biological Sciences Research Council (BBSRC) award to authors FM, XH, and ARB (BB/X008428/1). JK and YK are funded by the National Natural Science Foundation of China (32171082, 32471136).

References

Aguado, L., Parkington, K. B., Dieguez-Risco, T., Hinojosa, J. A., & Itier, R. J. (2019). Joint modulation of facial expression processing by contextual congruency and task demands. *Brain Sciences*, *9*(5), 116. https://doi.org/10.3390/brainsci9050116

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*(6835), 305-309. https://doi.org/10.1038/35077083

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M. & Lee, K. (2013). Development of own-race biases. *Visual Cognition, 21*(9), 1165-1182. https://doi.org/10.1080/13506285.2013.821428

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology, 17*(1), 35-55. https://doi.org/10.1080/026432900380472

Blau, V. C., Maurer, U., Tottenham, N., & McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions, 3*, 1-13. https://doi.org/10.1186/1744-9081-3-7

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/9780470743386

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Calvo, M. G., Beltrán, D., & Fernández-Martín, A. (2014). Processing of facial expressions in peripheral vision: Neurophysiological evidence. *Biological Psychology, 100*, 60-70. https://doi.org/10.1016/j.biopsycho.2014.05.007

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods, 44*, 1255-1265. https://doi.org/10.3758/s13428-012-0186-0

Carlson, J. M., & Reinke, K. S. (2010). Spatial attention-related modulation of the N170 by backward masked fearful faces. *Brain and Cognition, 73*(1), 20-27. https://doi.org/10.1016/j.bandc.2010.01.007

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. https://doi.org/10.1016/j.neuroimage.2021.118712

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology, 56*(11), e13437. https://doi.org/10.1111/psyp.13437

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Eimer, M. (2011). The face-sensitive N170 component of the event-related potential. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 329-344). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.001.0001

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *NeuroReport, 13*, 427-431. https://doi.org/10.1097/00001756-200203250-00013

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31. https://doi.org/10.1016/j.neuropsychologia.2006.04.022

Eimer, M., Holmes, A., & McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, and Behavioral Neuroscience, 3*(2), 97-110. https://doi.org/10.3758/CABN.3.2.97

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting Psychologists Press.

Faja, S., Dawson, G., Aylward, E., Wijsman, E. M., & Webb, S. J. (2016). Early event-related potentials to emotional faces differ for adults with autism spectrum disorder and by serotonin transporter genotype. *Clinical Neurophysiology, 127*(6), 2436-2447. https://doi.org/10.1016/j.clinph.2016.02.022

Fisher, K., Towler, J., & Eimer, M. (2017). Face identity matching is selectively impaired in developmental prosopagnosia. *Cortex, 89*, 11-27. https://doi.org/10.1016/j.cortex.2017.01.003

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion*, *14*(1), 61-92. https://doi.org/10.1080/026999300378996

Garrett-Ruffin, S., Hindash, A. C., Kaczkurkin, A. N., Mears, R. P., Morales, S., Paul, K., ... & 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

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1-9. https://doi.org/10.1038/sdata.2016.44

Herrmann, M. J., Aranda, D., Ellgring, H., Mueller, T. J., Strik, W. K., Heidrich, A., & Fallgatter, A. J. (2002). Face-specific event-related potential in humans is independent from facial expression. *International Journal of Psychophysiology, 45*(3), 241-244. https://doi.org/10.1016/S0167-8760(02)00033-8

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 55*, 498-509. https://doi.org/10.1016/j.neubiorev.2015.06.002

Holmes, A., Kiss, M., & Eimer, M. (2006). Attention modulates the processing of emotional expression triggered by foveal faces. *Neuroscience Letters*, *394*(1), 48-52. https://doi.org/10.1016/j.neulet.2005.10.002

Holmes, A., Vuilleumier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Cognitive Brain Research, 16*(2), 174-184. https://doi.org/10.1016/S0926-6410(02)00268-9

Ito, T. A., Larsen, J. T., Smith, N. K., & Caciopp, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*, 887-900. https://doi.org/10.1037//0022-3514.75.4.887

Jeffreys, H. (1961). *Theory of probability.* Oxford University Press.

Jovanović, V., Petrušić, I., Ković, V., & Savić, A. M. (2025). The practical implications of re-referencing in ERP studies: The case of N400 in the picture-word verification task. *Diagnostics, 15*(2), 156. https://doi.org/10.3390/diagnostics15020156

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: The effect of reference electrode site. *Clinical Neurophysiology, 116*(11), 2613-2631. https://doi.org/10.1016/j.clinph.2005.07.005

Kiss, M., & Eimer, M. (2008). ERPs reveal subliminal processing of fearful faces. *Psychophysiology, 45*(2), 318-326. https://doi.org/10.1111/j.1469-8986.2007.00634.x

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *European Journal of Neuroscience, 54*, 8406-8420. https://doi.org/10.1111/ejn.14992

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision, 9*(10), 6. https://doi.org/10.1167/9.10.6

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75-82. https://doi.org/10.1016/j.tics.2004.12.004

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Leppänen, J. M., Hietanen, J. K., & Koskinen, K. (2008). Differential early ERPs to fearful versus neutral facial expressions: A response to the salience of the eyes? *Biological Psychology, 78*(2), 150-158. https://doi.org/10.1016/j.biopsycho.2008.02.002

Leppänen, J. M., Kauppinen, P., Peltola, M. J., & Hietanen, J. K. (2007). Differential electrocortical responses to increasing intensities of fearful and happy emotional expressions. *Brain Research, 1166*, 103-109. https://doi.org/10.1016/j.brainres.2007.06.060

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146-157. https://doi.org/10.1111/psyp.12639

Luo, W., Feng, W., He, W., Wang, N. Y., & Luo, Y. J. (2010). Three stages of facial expression processing: ERP study with rapid serial visual presentation. *NeuroImage*, *49*(2), 1857-1867. https://doi.org/10.1016/j.neuroimage.2009.09.018

Miller, G. A., Lutzenberger, W., & Elbert, T. (1991). The linked-reference issue in EEG and ERP recording. *Journal of Psychophysiology, 5*(3), 273-276.

Niedenthal, P. M., & Ric, F. (2017). *Psychology of emotion* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781315276229

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG* (2nd ed.). Oxford University Press.

Ollikainen, J. O., Vauhkonen, M., Karjalainen, P. A. & Kaipio, J. P. (2000). Effects of electrode properties on EEG measurements and a related inverse problem. *Medical Engineering and Physics, 22*(8), 535-545. https://doi.org/10.1016/S1350-4533(00)00070-9

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, *30*(6), 808-822. https://doi.org/10.1016/j.neubiorev.2006.06.007

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331. https://doi.org/10.1037/0033-295X.97.3.315

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466-478. https://doi.org/10.1037/0096-3445.130.3.466

Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2012). AMICA: An adaptive mixture of independent component analyzers with shared components. *Swartz Center for Computational Neuroscience, University of California San Diego, Tech. Rep*, *35*.

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, C. S. Y., Beste, C., … & 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

Pegna, A. J., Landis, T., & Khateb, A. (2008). Electrophysiological evidence for early non-conscious processing of fearful facial expressions. *International Journal of Psychophysiology*, *70*(2), 127-136. https://doi.org/10.1016/j.ijpsycho.2008.08.007

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*, 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data, 6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148-158. https://doi.org/10.1038/nrn2317

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), 773-783. https://doi.org/10.1038/nrn2920

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11458-11463. https://doi.org/10.1073/pnas.172403899

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181-197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. https://doi.org/10.1037/0033-2909.124.3.372

Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine.* Elsevier. https://doi.org/10.1016/0166-2236(89)90083-0

Rellecke, J., Palazova, M., Sommer, W., & Schacht, A. (2011). On the automaticity of emotion processing in words and faces: Event-related brain potentials evidence from a superficial task. *Brain and Cognition*, *77*(1), 23-32. https://doi.org/10.1016/j.bandc.2011.07.001

Rellecke, J., Sommer, W., & Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology, 90*(1), 23-32. https://doi.org/10.1016/j.biopsycho.2012.02.002

Rellecke, J., Sommer, W., & Schacht, A. (2013). Emotion effects on the N170: A question of reference?. *Brain topography, 26*, 62-71. https://doi.org/10.1007/s10548-012-0261-y

Righart, R., & De Gelder, B. (2006). Context influences early perceptual analysis of faces — An electrophysiological study. *Cerebral Cortex, 16*(9), 1249-1257. https://doi.org/10.1093/cercor/bhj066

Righart, R., & De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience, 3*(3), 270-278. https://doi.org/10.1093/scan/nsn021

Rossion, B., & Jacques, C. (2011). The N170: Understanding the time course of face perception in the human brain. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of ERP components* (pp. 115-142). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.001.0001

Rossignol, M., Philippot, P., Douilliez, C., Crommelinck, M., & Campanella, S. (2005). The perception of fearful and happy facial expression is modulated by anxiety: an event-related potential study. *Neuroscience Letters, 377*(2), 115-120. https://doi.org/10.1016/j.neulet.2004.11.091

Royal Society of Chemistry. (2020). *Joint commitment for action on inclusion and diversity in publishing*. Retrieved July 13, 2025, from https://www.rsc.org/policy-and-campaigning/science-culture/joint-commitment-for-action-on-inclusion-and-diversity-in-publishing

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology. 39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Schindler, S., Bruchmann, M., Bublatzky, F., & Straube, T. (2019). Modulation of face- and emotion-selective ERPs by the three most common types of face image manipulations. *Social Cognitive and Affective Neuroscience, 14*(5), 493-503. https://doi.org/10.1093/scan/nsz027

Schindler, S., Bruchmann, M., Steinweg, A.-L., Moeck, R., & Straube, T. (2020). Attentional conditions differentially affect early, intermediate and late neural responses to fearful and neutral faces. *Social Cognitive and Affective Neuroscience, 15*(7), 765-774. https://doi.org/10.1093/scan/nsaa098

Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex, 130*, 362-386. https://doi.org/10.1016/j.cortex.2020.06.010

Schindler, S., Tirloni, C., Bruchmann, M., & Straube, T. (2021). Face and emotional expression processing under continuous perceptual load tasks: An ERP study. *Biological Psychology, 161*, 108056. https://doi.org/10.1016/j.biopsycho.2021.108056

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: Event-related brain potential studies. *Progress in Brain Research*, *156*, 31-51. https://doi.org/10.1016/S0079-6123(06)56002-9

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion*, *4*(2), 189. https://doi.org/10.1037/1528-3542.4.2.189

Smith, N. K., Cacioppo, J. T., Larsen, J. T., & Chartrand, T. L. (2003). May I have your attention please: Electrocortical responses to positive and negative stimuli. *Neuropsychologia, 41*, 171-183. https://doi.org/10.1016/S0028-3932(02)00147-1

Sprengelmeyer, R., & Jentzsch, I. (2006). Event related potentials and the perception of intensity in facial expressions. *Neuropsychologia, 44*(14), 2899-2906. https://doi.org/10.1016/j.neuropsychologia.2006.06.020

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, *30*(3), 829-841. https://doi.org/10.1016/s0896-6273(01)00328-2

Weidner, E. M., Schindler, S., Grewe, P., Moratti, S., Bien, C. G., & Kissler, J. (2022). Emotion and attention in face processing: Complementary evidence from surface event-related potentials and intracranial amygdala recordings. *Biological Psychology*, *173*, 108399. https://doi.org/10.1016/j.biopsycho.2022.108399

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for ANOVA designs. *The American Statistician, 66*(2), 104-111. https://doi.org/10.1080/00031305.2012.695956

Figure 1. Stimulus display and trial sequence of the current study. Each trial starts with a 300-ms display containing two identical faces (external features removed) and two line segments laterally presented. A 2,000-ms interval (fixation only) follows before the next trial. In the emotion task, participants will be asked to judge whether the faces are emotional or neutral. In the line task, participants will judge whether the line segments have the same length.

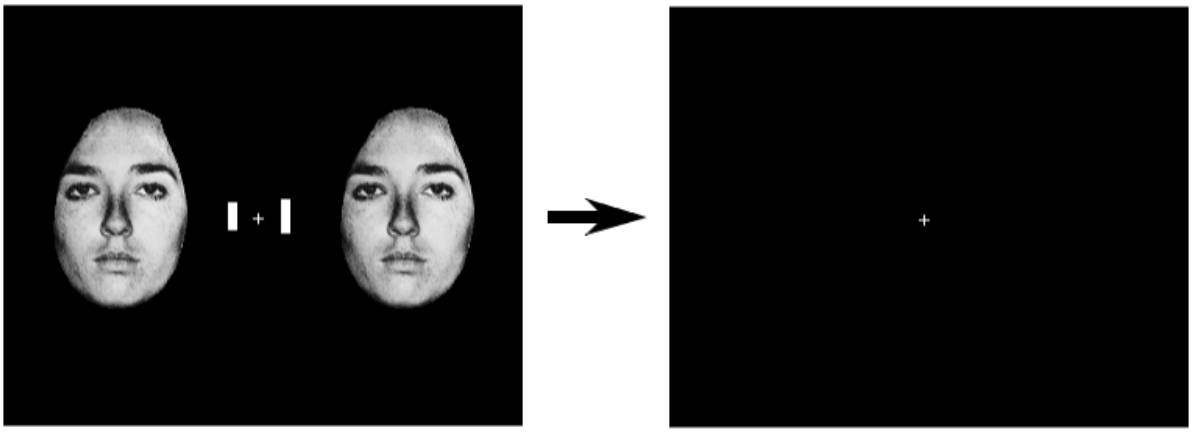

Table 1. EEG system setups of the replicating labs.

Replicating lab	**EEG system (brand and model, highest sampling rate, lowest high-pass filter setting)**	**Electrodes**	**Impedance/offset threshold**	**Reference & Ground**	**Line noise frequency (Hz)**	**Typical artefact rejection threshold (µV)**	**Screen**	**Stimulus presentation**	**Viewing distance (cm)** **To be adjusted to allow 100% scaling**	**Image scaling factor**
Bournemouth University	Biosemi ActiveTwo Mk2 2048 Hz DC capable	Pin- and flat-type active electrodes, Ag/AgCl 64 scalp, HEOG, earlobes	Offset < 40 mV	REF: none GND: CMS/DRL	50	50 - 80	HP EliteDisplay E231 60 Hz 1920 x 1080 51.0 x 28.7 cm	PsychoPy 2023.1.2	63.5	100%

Koç University	Brain Vision actiCHamp Plus 100 kHz DC capable	actiCap active electrodes Ag/AgCl 28 scalp, 2 HEOG, 2 earlobe	Impedance < 10 kΩ	REF: Right Earlobe GND: Fpz	50	n/a (semi-automatic with step function convolution for EOG, manual approach for EEG)	LG Flatron M2280DM 60 Hz 1920 x 1080 47.6 x 26.8 cm	PsychoPy 2023.2.3	59.2	100%
Western Sydney University	Biosemi ActiveTwo Mk2 2048 Hz DC capable	Pin- and flat-type active electrodes, Ag/AgCl 64 scalp, HEOG	Offset < 40 mV	REF: none GND: CMS/DRL	50	n/a	VIEWPixx /EEG 120 Hz 1920 x 1080 53.2 x 29.9 cm	PsychoPy 2023.2.3	66.2	100%
Manchester Metropolitan University	Ant Neuro eego mylab 16 kHz	Waveguard Nets	Impedance < 30 kΩ	REF: Cz and CPz GND: Mastoids	50	300 (peak to peak)	BenQ GL2760-T 60 Hz	PsychoPy 2025.1.1	74.3	100%

	DC capable	(64-channel nets)					1920 x 1080 59.7 x 33.5 cm			
Brandeis University	Brainvision BrainAmp Standard 5kHz/channel (160kHz total) DC capable	Brainvision actiCAP Ag/AgCl 60 scalp, HEOG, 2 Earlobe (note: change to actiCAP slim/snap in September)	Impedance < 10 kΩ	REF: FCz GND: AFz	60	80 - 100	Alienware AW2724HF 60 Hz 1920 x 1080 59.7 x 33.5 cm	PsychoPy 2025.1.1	74.3	100%
Sun Yat-sen University	Neuroscan SynAmps 2 Model	Neuroscan SynAmps	Impedance < 10 kΩ	REF: Mastoids Online: M1,	50	100 - 150	FangZheng FC2453D, 60 Hz	PsychoPy 2023.2.3	66.2	100%

	8050 (max 20kHz) DC capable	2 Ag/AgCl 64 scalp 2 HEOG, 2 VEOG, 2 Mastoid		Offline: (M1+M2)/ 2 GND: Fpz			1920 x 1080 53.2 × 29.9 cm			
Universidad Complutense de Madrid	BrainProducts actiCHamp Plus 50 KHz (64 channels)/ 100 kHz (32 channels) DC capable	actiCap active electrodes Ag/AgCl 64 scalp, 1 HEOG, 1 VEOG	Impedance < 10 kΩ	REF: FCz (adjustable in software) GND: Fpz	50	80 - 100	Dell OptiPlex 7780 60 Hz 1920 x 1080 60 x 33.5 cm	PsychoPy 2025.1.1	74.7	100%

Table 2. EEG data acquisition settings and pre-processing steps of the original, replication, and advanced pipelines.

Stages	**Settings/Steps**	**Eimer et al. (2003)**	**Replication pipeline**	**Advanced pipeline**
EEG acquisition	***Online reference***	Left earlobe	Single-electrode reference per each lab setting (single earlobe preferred)	
	Online filter	0.1-40 Hz bandpass	None (0.1 Hz highpass if DC recording is not possible)	
	Sampling rate	200 Hz	500 or 512 Hz	
EEG pre-processing	***Re-referencing***	Averaged earlobes	Averaged earlobes or mastoids. If earlobes/mastoids are not possible, TP9/TP10 will be used.	Averaged earlobes or mastoids. If earlobes/mastoids are not possible, TP9/TP10 will be used.
	Resampling	None	Down-sampling to 200 Hz	None
	Offline filter	None	0.1-40 Hz bandpass (40 Hz lowpass if a 0.1-Hz highpass filter is applied during data acquisition).	0.1-40 Hz bandpass (40 Hz lowpass if a 0.1-Hz highpass filter is applied during data acquisition).
	ICA	None	None	AMICA on 2-Hz highpass filtered data. ICs identified by

				ICLabel as eye, muscle, heart, line noise, and channel noise with ≥ 80% probability will be removed from the original unfiltered data.
	Segmentation	-100 to 700 ms	-100 to 700 ms	-100 to 700 ms
	Baseline correction	-100 to 0 ms	-100 to 0 ms	-100 to 0 ms
	Artefact rejection	HEOG (bipolar) exceeding ±30 µV Fpz exceeding ±60 µV Any electrode exceeding ±80 µV	HEOG (bipolar) exceeding ±30 µV Fpz (Fp1 if Fpz unavailable) exceeding ±60 µV Any key electrode (F3, Fz, F4, C3, Cz, C4, P7, P3, Pz, P4, P8, PO7, PO8) exceeding ±80 µV	HEOG (bipolar) exceeding ±30 µV Fpz (Fp1 if Fpz unavailable) exceeding ±60 µV Any key electrode (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, P7, P8, PO7, PO8) exceeding ±80 µV Any key electrode (F3, Fz, F4, C3, Cz, C4, P7, P3, Pz, P4, P8, PO7, PO8) showing absolute activity below 0.5 µV for more than 200 ms

	Participant removal	Participant removal if excessive eyeblinks	Participant removal if more than 50% of trials are rejected for any of the two tasks	Participant removal if more than 50% of trials are rejected for any of the two tasks
	Averaging	Only trials with correct behavioural responses	Only trials with correct behavioural responses	Only trials with correct behavioural responses
ERP measurement	***Mean amplitude quantification***	Visual inspection based on maximal activities.	Channels and time windows as in the original study.	Channels as in the original study. For each lab, time windows are determined by collapsed localiser for each electrode group (frontal, central, parietal, temporal, lateral occipital). In the localiser waveform for each analysis, time window boundaries are set at 50% of the peak amplitude, which is determined in the time windows used for mean amplitude quantification in the original study.

Table 3. Planned ERP data that will enter the meta-analyses corresponding to the replication's hypotheses. Both frequentist and Bayesian analyses will be performed on the data from each of the two EEG pre-processing pipelines. In the frequentist analyses, standardised effect size (Cohen's d) and its SE from each analysis will be derived from each lab and put into a frequentist meta-analysis. In the Bayesian analyses, standardised effect sizes (Cohen's d) and their associated uncertainty (credible intervals) from each individual dataset will be extracted and entered into a Bayesian random-effects meta-analysis. In case the frequentist and Bayesian analysis results diverge, we will prioritise the frequentist interpretation complemented by the Bayesian results.

Task	**ERP measure**	**Hypothesis**	**Analysis**	**Bayesian prior**	**Predicted result**
Emotion	Early frontal positivity	H1.1	ANOVA [a]	0.707 [b]	Emotion [c] + [d]
	Frontal/central positivity > 200 ms	H1.2	ANOVA	0.707	Emotion +
	Lateral temporal/occipital negativity > 200 ms	H1.3	ANOVA	0.707	Emotion +
	Frontal/central/parietal positivity > 300 ms	H1.4	ANOVA	0.707	Emotion +
	Frontal/central/parietal positivity > 500 ms	H1.5	ANOVA	0.707	Emotion +
	N170	H3	ANOVA	0.3325, 0.2062, 0.4219 [e]	Emotion - [f]
		H3	*t*-test for each Block Type	0.3325, 0.2062, 0.4219	Emotion -
	All the six activities listed above	H4	Pairwise *t*-tests across emotions (emotional trials only) [g]	0.707	Emotion Type [h] -
Line	Early frontal positivity	H2.1	ANOVA	0.707	Emotion -
	Frontal/central positivity > 200 ms	H2.2	ANOVA	0.707	Emotion -

	Lateral temporal/occipital negativity > 200 ms	H2.3	ANOVA	0.707	Emotion -
	Frontal/central/parietal positivity > 300 ms	H2.4	ANOVA	0.707	Emotion -
	Frontal/central/parietal positivity > 500 ms	H2.5	ANOVA	0.707	Emotion -
	N170	H3	ANOVA	0.3325, 0.2062, 0.4219	Emotion -
	All the six activities listed above	H4	Pairwise *t*-tests across emotions (emotional trials only)	0.707	Emotion Type -

[a] Two-way (Emotion x Block Type) repeated-measures ANOVA.

[b] The listed priors are set for the effects specified in the "Predicted result" column. All unspecified priors will use 0.707.

[c] Emotion effect = emotional vs. neutral.

[d] The symbol '+' denotes results of $p < .02$ or $BF_{10} \geq 3$.

[e] The three Emotion effect priors are based on a meta-analysis (Hinojosa et al., 2015), targeting the mean, the smallest, and the largest effect sizes, respectively. All three priors will be used in separate meta-analyses for robustness assessment.

[f] The symbol '-' denotes results of $p > .02$ or $BF_{10} \leq 0.333$.

[g] Only take place when $p < .02$ in the frequentist one-way ANOVA (emotional trials only, on the whole sample across all labs).

[h] Emotion Type effect = difference between two emotions. This will vary depending on the emotion pair in each *t*-test.

Appendix 1. Demographic questions and answer options about gender, ethnic origin (ancestry), and race (adapted from Joint Commitment for Action on Inclusion and Diversity in Publishing).

Please answer the questions below when being asked by researcher(s):

With which gender do you most identify? Please select one option:

- Woman
- Man
- Non-binary or Gender diverse
- Self-describe: _____
- Prefer not to disclose

What are your ethnic origins or ancestry? Please select ALL the geographic areas from which your family's ancestors first originated:

- Western Europe (e.g., Greece, Sweden, United Kingdom)
- Eastern Europe (e.g., Hungary, Poland, Russia)
- North Africa (e.g., Egypt, Morocco, Sudan)
- Sub-Saharan Africa (e.g., Kenya, Nigeria, South Africa)
- West Asia / Middle East (e.g., Iran, Israel, Saudi Arabia)
- South and Southeast Asia (e.g., India, Indonesia, Singapore)
- East and Central Asia (e.g., China, Japan, Uzbekistan)
- Pacific / Oceania (e.g., Australia, Fiji, Papua New Guinea)
- North America (Canada, United States)
- Central America and Caribbean (e.g., Jamaica, Mexico, Panama)
- South America (e.g., Brazil, Chile, Colombia)
- Self describe: _____

How would you identify yourself in terms of race? Please select ALL the groups that apply to you:

- Asian or Pacific Islander
- Black
- Hispanic or Latino/a/x
- Indigenous (e.g., North American Indian Navajo, South American Indian Quechua, Aboriginal or Torres Strait Islander)
- Middle Eastern or North African
- White
- Self describe: _____