

A multilab investigation into the N2pc as an indicator of attentional selectivity: Direct replication of Eimer (1996)

Martin Constant^{1,2,*}, Ananya Mandal^{2,3}, Dariusz Asanowicz⁴, Motonori Yamaguchi⁵, Helge Gillmeister⁵, Dirk Kerzel⁶, David Luque⁷, Francesca Pesciarelli⁸, Thorsten Fehr¹, Faisal Mushtaq⁹, Yuri G. Pavlov^{10, 11}, and Heinrich R. Liesefeld^{1,2,*}

¹Institute of Psychology, University of Bremen, Germany

²Graduate School of Systemic Neuroscience, Ludwig-Maximilians-Universität München, Germany

³Department of Psychology, Ludwig-Maximilians-Universität München, Germany

⁴Institute of Psychology, Jagiellonian University, Kraków, Poland

⁵Department of Psychology, University of Essex, United Kingdom

⁶Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, Switzerland

⁷School of Psychology, Universidad de Málaga, Spain

⁸Department of Biomedical, Metabolic, and Neural Sciences, University of Modena and Reggio Emilia, Italy

⁹School of Psychology, University of Leeds, United Kingdom

¹⁰Department of Psychology, Ural Federal University, Yekaterinburg, Russia

¹¹Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany

Abstract

The N2pc is widely employed as an electrophysiological marker of an attention allocation. This interpretation was in no small part driven by the observation of an N2pc elicited by an isolated relevant target object, which was reported as Experiment 2 in Eimer (1996; The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, 99, 225–234). All subsequent refined interpretations of the N2pc had to take this crucial finding into account. Despite its central role for neurocognitive attention research, there have been no direct replications and only few conceptual replications of this seminal work. Within the context of #EEGManyLabs, an international community-driven effort to replicate the most influential EEG studies ever published, the present study was selected due to its impact on the study of selective attention. We propose to provide a high-powered direct replication, carefully following all the steps laid out in Pavlov et al. (2021; #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex*, 144, 213–229) to assure a high-quality direct replication of the original study. In doing so, we expect to shed further light on the functional significance of the N2pc as an electrophysiological marker of attentional selectivity.

Keywords: N2pc, spatial attention, visual attention, replication, #EEGManyLabs

The N2pc is a component of the lateralized event-related potential evoked by a stimulus presented in one visual hemifield, which – due to the physiology of the visual system – is first processed in brain areas contralateral to the presentation side. The N2pc usually expresses as a transient negativity in the difference wave between activity measured at parieto-occipital electrodes contra- *minus* ipsilateral to the presentation of the stimulus in question. It typically starts around 200 ms after stimulus onset and rises and falls within around 150 ms with systematic variations in timing due to task manipulations (Liesefeld et al., 2017; Luck, 2012; Luck & Hillyard, 1990; Töllner et al., 2011).

The N2pc is most often used as a marker of shifts of attention. From observing it, numerous studies conclude that

the lateralized stimulus was attentionally processed (e.g., Burra & Kerzel, 2013; Eimer & Kiss, 2008; Hickey et al., 2006; Lien et al., 2008; Töllner et al., 2012; Woodman & Luck, 1999). This interpretation of the N2pc component was sparked by the seminal work of Eimer (1996), which is the target study we attempt to replicate here.

Our replication study is situated within the context of a large community-driven international project, #EEGManyLabs, whose ambition is to run high-powered replications of many influential EEG studies through multi-lab collaborations. The present study was selected as a target for replication by an international group of EEG experts based on its scientific impact (see Pavlov et al., 2021, for details on the selection procedure).

All participants in the present replication project volunteered because (a) they use or plan to use the N2pc in their work and (b) they agreed that Eimer (1996) had a strong influence on popularizing the N2pc component as a tool in attention research and on popularizing the particular interpretation of the N2pc as an electrophysiological correlate of a candidate target stimulus' selection (Eimer, 2014). For these reasons, replicating this particular study seems of utmost importance for neurocognitive research on selective attention.

Crucially, the researchers who first discovered the N2pc (Luck & Hillyard, 1990) interpreted it not as reflecting an attention allocation to the relevant stimulus, but rather as reflecting the suppression of the display elements surrounding the relevant stimulus (Luck & Hillyard, 1994; Luck et al., 1993). On that background, Eimer (1996) demonstrated that the N2pc emerges even if there are no elements surrounding the relevant stimulus, but only a single irrelevant stimulus is presented on the other side of the display (which had the sole purpose of balancing visual stimulation).

Eimer (1996)'s finding does not exclude alternative interpretations of the N2pc brought forward subsequently, which suggest that the typically observed N2pc might be a composite reflecting both enhancement of the relevant stimulus and suppression of the irrelevant stimulus on the opposite side (Hickey et al., 2009 – which is also the most notable conceptual replication apart from the two other experiments reported in the original paper), or that the N2pc reflects engagement at the location of the relevant stimulus rather than the shift of attention proper (Zivony et al., 2018). It is also possi-

ble that the N2pc reflects some kind of ambiguity resolution in favor of the target that is required due to the presence of other display elements even if this is only a single irrelevant item on the opposite display side (Luck, 2012; Luck et al., 1997). In any case, Eimer's 1996 finding of an N2pc to a non-surrounded relevant stimulus was undeniably influential in triggering discussions about the functional significance of the N2pc and must be accounted for in any serious speculation on what cognitive process the N2pc reflects. Even though, over the decades following the publication of Eimer (1996), the N2pc has been used extensively as a marker of the allocation of spatial attention towards a particular stimulus (*attention allocation*), only few N2pc studies have presented the relevant stimulus without surrounding elements (Hickey et al., 2009; van Moorselaar & Slagter, 2019).

The existence of an N2pc in the study by Eimer (1996) was supported by an effect of laterality in the predetermined time window 220 – 300 ms after display onset that was used throughout three experiments. In the most crucial Experiment 2 that we aim to replicate here, N2pcs were tested and observed in two conditions: with the relevant and irrelevant object being (a) forms or (b) color patches. The task was to discriminate whether an M or a W was shown or whether a color patch was green or blue, respectively, with the respective irrelevant stimuli being a collection of vertical lines or a yellow patch (see Fig. 1c-d). Thus, we aim to replicate the two N2pcs observed in Experiment 2 of Eimer (1996; see Fig. 1a-b).

Beyond these main effects of interest, a serendipitous finding is worth mentioning here: The N2pc to forms appeared to be larger in amplitude and in additional time windows compared to the N2pc for color patches. Eimer (1996) interpreted the amplitude effect as a consequence of the higher difficulty of discriminating the M and W compared to discriminating green and blue. Thus, we expect to replicate a higher amplitude for an N2pc elicited by forms compared to color patches (see Fig. 1b).

Methods

Transparency and openness statement

We report how we determine our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria are established prior to data analysis, all manipulations, and all measures in the study.

Stimuli, procedure & design

The experiment was developed in OpenSesame 3.3.14 (Mathôt et al., 2012) with the PsychoPy (Peirce et al., 2019) backend used for stimulus presentation and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) for timings and response collection (the Python environment file

We have no known conflict of interest to disclose.

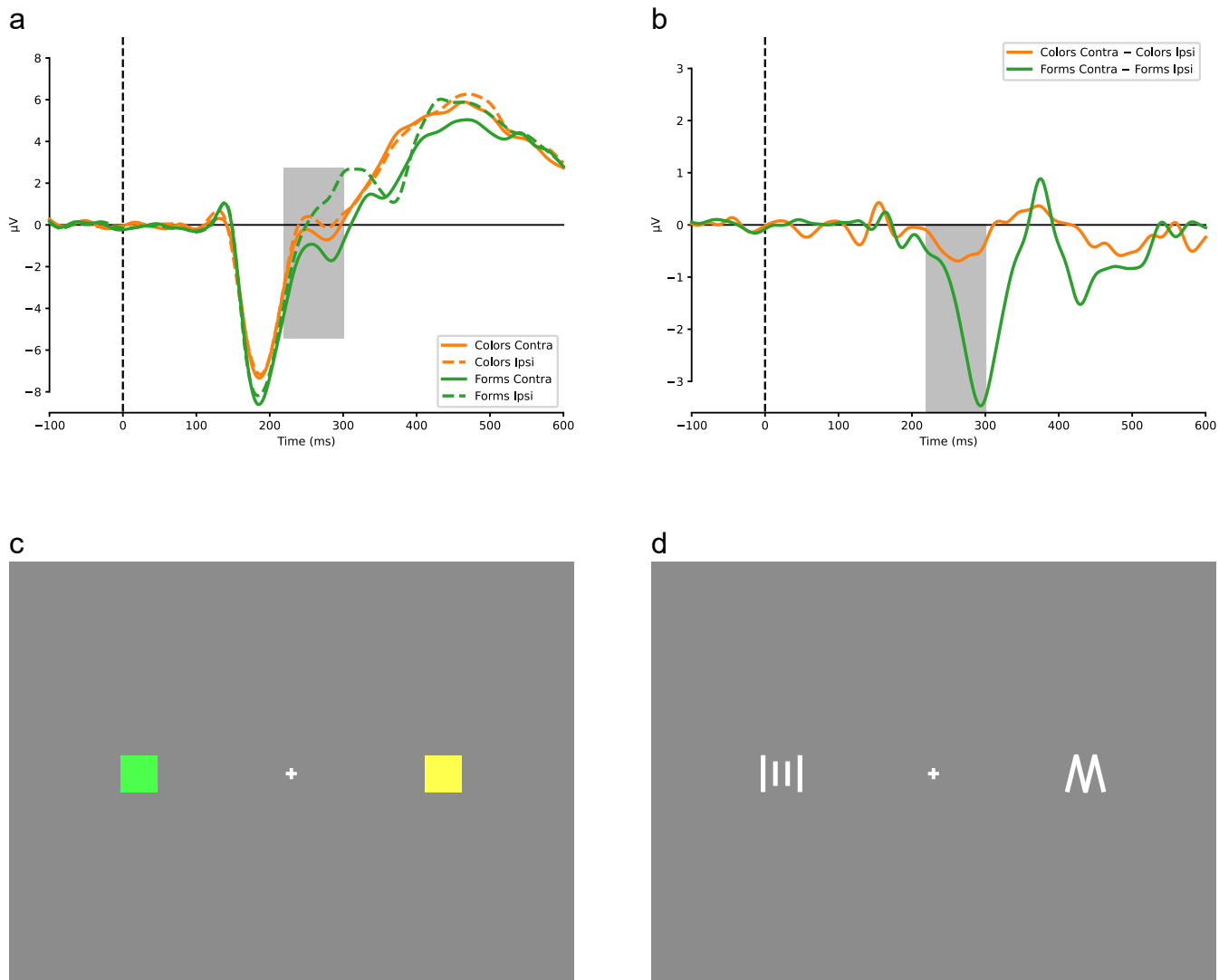
-  [Martin Constant](#)
-  [Ananya Mandal](#)
-  [Dariusz Asanowicz](#)
-  [Motonori Yamaguchi](#)
-  [Helge Gillmeister](#)
-  [Dirk Kerzel](#)
-  [David Luque](#)
-  [Francesca Pesciarelli](#)
-  [Thorsten Fehr](#)
-  [Faisal Mushtaq](#)
-  [Yuri G. Pavlov](#)
-  [Heinrich R. Liesefeld](#)

*Correspondence concerning this article should be addressed to Martin Constant, Email: martin.constant@uni-bremen.de or Heinrich R. Liesefeld, Email: heinrich.liesefeld@uni-bremen.de

#EEGManyLabs is funded by the DFG (PA 4005/1-1), provided to Y.G.P. Lead Replicating Senior Author (H.R.L) and First Author are funded by the DFG (LI 2868/3-1). A.M. is funded by the DFG (RTG 2175). D.A. is funded by the National Science Center of Poland (2016/22/E/HS6/00139). D.L. is funded by the MCIN/AEI (PID2021-126767NB-I00) and the Junta de Andalucía (PROYEXCEL_00287). The funders had no role in study design or the decision to submit the work for publication.

Figure 1

Reconstructed ERPs and displays of the experiment.



Note. (a) and (b) The ERPs were digitized from the original manuscript with Engauge (Mitchell et al., 2019), interpolated to 1000 Hz using CubicSpline interpolation with scipy v1.10.0 (Virtanen et al., 2020), then low-passed filtered at 30 Hz (passband edge; one-pass, zero-phase, non-causal FIR filter, Hamming-windowed sinc, filter order 440) with MNE version 1.3.0 (Gramfort et al., 2013), visualization was also created with MNE. The shaded area represents the original analysis time window (220 – 300 ms). (c) and (d) Search displays were recreated in OpenSesame using information from the original study’s manuscript and personal communication with the author. Panel (c) represents the *colors* condition and panel (d) represents the *forms* condition. A version of this figure with inverted Y axis is available on the OSF repository.

and the experiment are provided on <https://osf.io/4ux8r/>. The color values we use were obtained from personal communication with the original author and reflect his best guess. A standard operating protocol notably including how to set up and run the experiment is provided on the OSF repository (<https://osf.io/4ux8r/wiki>).

A 100% white central fixation cross (line length: 0.24 degrees of visual angle [dva; assuming that the viewing dis-

tance indicated in the experimental settings is maintained], line width: 0.04 dva) is displayed against a 55% gray background for the whole experiment (i.e., it only disappears during breaks). In half of the experimental blocks (*form discrimination*), a letter stimulus (M or W, line width: 0.08 dva) is presented together with either the same letter (*target-only arrays*) or a distractor (*distractor arrays*) which is an arrangement of two long and two short vertical bars (line width: 0.08

dva). In the other experimental half (color discrimination), one square in a target color (blue [RGB: 30%, 30%, 100%] or green [RGB: 30%, 100%, 30%]) is presented together with a square of the same color (target-only arrays) or a distractor (distractor arrays) which is a yellow square (RGB: 100%, 100%, 30%). In each trial, the two stimuli appear 3.3 dva to the right and left of the center of the screen for 150 ms; each stimulus subtends 0.8×0.8 dva. From the onset of the stimulus array until 2000 ms after its disappearance (i.e., 2150 ms after onset), participants must indicate which target (M or W; blue or green) they see by pressing the left or right key of their response device, independently of the target's side. The response-key assignment is counterbalanced across participants. Keypresses are stored in an asynchronous buffer. After 2150 ms this buffer is read and the first key pressed (if any) is considered to be the participant's response. Timeouts (i.e., no key pressed) are considered as errors.

Each participant starts with one condition (M vs. W or blue vs. green; order counterbalanced) and performs 6 blocks of 66 trials of this condition before switching to the other condition with the same number of trials. There are 4 distractor-array configurations (target identity [2] \times target side [2]) and there are 2 configurations for target-only arrays (target identity [2]). Each of these 6 conditions is presented an equal number of times in a block (11 times per block).

Participants are instructed not to move their eyes from the fixation cross. To train them not to move their eyes, a practice block runs until the experimenter judges from the HEOG waves that participants are holding their eyes sufficiently still.

Note that artifacts induced by horizontal eye movements are of particular relevance in N2pc studies, because gaze is likely to be directed at the lateralized stimulus for which attention allocations are examined (here: the target) and would therefore produce lateralized activity that confounds the lateralized activity of interest. Furthermore, an eye movement towards the target would center the image of the target on the retina and thereby invalidate the reasoning behind the lateralized presentation.

This practice block also serves as training to learn the response-key assignments and, therefore immediate feedback is provided. In particular, in the event of an incorrect response, a large gray "X" is displayed for 500 ms between two practice trials and in the event of a timeout, a gray hourglass is presented for the same duration. Correct responses do not prompt the appearance of any feedback, the fixation cross simply remains for an extra 500 ms.

EEG data acquisition

Each laboratory is vetted by the corresponding authors. More precisely, a video of the experimental setup is sent to the corresponding authors as well as a pilot dataset to standardize the data acquisition process as much as possi-

ble. More laboratories might join after the in-principle acceptance. The setup of each lab is described in [Table 1](#).

EEG offline preprocessing

The EEG data is preprocessed with two slightly different pipelines and results are extracted with two different methods from each pipeline, resulting in four pipeline combinations. The first "Original" pipeline is the direct replication attempt, and the alternative pipelines will be used to cross-validate the results with more modern processing techniques.

Original pipeline

The first pipeline aims to be as close as possible to the original pipeline and is therefore called the "Original" pipeline. It goes as follows:

EEG data is imported from the original recording format to EEGLAB (with the most recent stable version available at the time of data analysis; Delorme and Makeig, 2004). After import, the markers are cleaned and harmonized to a common scheme, and markers reflecting the reaction time are added from information contained in the behavioral file. At this point, only for flatline (channel blocking) detection, a copy of the dataset is created and high-passed filtered at 1 Hz (bandpass edge) with "pop_eegfiltnew(EEG, 'locutoff', 1, 'usefftfilter', 1)" and with periods of data where no marker was sent for more than 5000 ms removed. If a mastoid electrode or PO7 or PO8 is flat (absolute voltage $< 4.5e-15\mu\text{V}$) for more than 30 seconds in this copied dataset, the participant is excluded and further processing is not performed. Next, the electrode layout in the original data set is harmonized (i.e., referenced to the BESA template) and data is re-referenced to the average of the mastoids. Data is then high-pass filtered at 0.1 Hz (bandpass edge; -6 dB cutoff at 0.05 Hz) using the "pop_eegfiltnew(EEG, 'locutoff', 0.1, 'usefftfilter', 1)" function from EEGLAB (one-pass, zero-phase, non-causal FIR filter, Hamming-windowed sinc, filter order depending on acquisition sampling rate), and then low-pass filtered at 40 Hz (bandpass edge; -6 dB cutoff at 45 Hz) using "pop_eegfiltnew(EEG, 'hicutoff', 40, 'usefftfilter', 0)". Finally, data is downsampled to 200 Hz. These filters and downsampling were designed to mimic the original study's amplifier recording settings.

Then, epochs of -100ms to 600ms relative to the onset of the display are created (baseline correction: -100ms - 0 ms). Only epochs for distractor arrays where the participant's response was correct are created. A bipolar horizontal EOG channel is created by subtracting the right HEOG from the left HEOG and a bipolar vertical EOG channel is created by subtracting the inferior VEOG from the superior VEOG (or Fp2 if no dedicated superior VEOG was recorded). Note that in the original study, due to the low number of available channels at the time, no inferior VEOG was recorded and, instead, the right HEOG was used. Epochs with an absolute voltage

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

Participating university	Manufacturer Amplifier Sampling rate	Electrodes Impedance threshold	Reference Ground	Hardware filters	EEG PC OS Recording software (version)	Line noise frequency	Screen	Display PC OS	Compensation
LMU München	BrainProducts BrainAmp DC 1000 Hz	Ag/AgCl ActiCap Slim (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 15 k Ω	REF: FCz GND: Fpz	HP: 0.016 Hz 1st order 6 dB/octave LP: 250 Hz 5th order Butterworth 30 dB/octave	Windows XP BrainVision Recorder (v1.20.0601)	50 Hz	VIEWPixx/3D (120Hz, scanning backlight)	Windows 10	Course credits or 10 €/h
Jagiellonian University	BioSemi ActiveTwo Mk2 1024 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 10 k Ω	REF: CMS GND: DRL	HP: DC LP: 200 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v7)	50 Hz	21" LCD monitor (60 Hz, 1920x1080)	Windows 10	50 zł/h
University of Essex	Compumedics Neuroscan SynAmps RT 1000 Hz	Ag/AgCl BrainCap (26 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 15 k Ω	REF: M1 GND: AFz	HP: 0.05 Hz 6 dB/octave LP: 100 Hz	Windows 10 Curry 8	50 Hz	Dell S2419HGF (120 Hz)	Windows 10	Course credits or 8 £/h
Université de Genève	BrainProducts actiCHamp 1000 Hz	Ag/AgCl ActiCap Slim (26 scalp + 2 HEOG + 2 VEOG + 2 earlobes) 10 k Ω	REF: Cz GND: AFz	HP: DC LP: 250 Hz	Windows 10 BrainVision Recorder (v1.25.0001)	50 Hz	VIEWPixx Lite (100Hz, normal backlight)	Windows 10	Course credits or 20 CHF/h
Universidad de Málaga	BrainProducts BrainAmp DC 1000 Hz	Ag/AgCl ActiCap Slim (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 15 k Ω	REF: FCz GND: Fpz	HP: 0.016 Hz 1st order 6 dB/octave LP: 250 Hz 5th order Butterworth 30 dB/octave	Windows 10 BrainVision Recorder (v1.24.0101)	50 Hz	17" LCD monitor	Windows 10	Course credits or 10 €/h
University of Modena and Reggio Emilia	BrainProducts actiCHamp Plus 1000 Hz	Ag/AgCl ActiCap Slim (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 20 k Ω	REF: FCz GND: Fpz	HP: DC LP: 250 Hz	Windows 10 BrainVision Recorder (v1.25.0101)	50 Hz	Philips 107B (60 Hz)	Windows 10	Course credits
Universität Bremen	TMS International REFA 512 Hz	Ag/AgCl ActiCap Slim (62 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 15 k Ω	REF: FCz GND: Cheek	HP: 0.001 Hz LP: 102.4 Hz 5th order FIR sinc filter	Windows XP eemagine (v3.3.02)	50 Hz	Belinea 1970 S1 (75 Hz)	Windows 7	Course credits

from the EOGs (non-bipolar), PO7 or PO8 below 1 μV for at least 350 contiguous milliseconds are rejected. Epochs are also rejected if the absolute amplitude of the bipolar VEOG is larger than 60 μV or if the absolute amplitude of the bipolar HEOG is larger than 25 μV at any timepoint in the epoch. The data is then averaged with ERPLAB (most recent stable version available at the time of data analysis; Lopez-Calderon & Luck, 2014). The left and right electrodes are then converted to contralateral or ipsilateral electrodes and contralateral *minus* ipsilateral difference waves are created. At this point, if the maximal absolute voltage of the HEOG difference wave, in the ERP calculated across all conditions, exceeds 2 μV at any time point, the participant is rejected from further analyses. The mean voltages for each collapsed condition (i.e., letters instead of separate M/W, colors instead of separate blue/green) and each side (ipsilateral or contralateral) from 220 to 300 ms are then extracted and statistically analyzed with paired-sample *t* tests (see Confirmatory analysis plan).

The paired-sample *t* test is performed with a custom implementation in MATLAB that requires the Statistics and Machine Learning Toolbox. In addition to the typical outputs (e.g., *t* value, *p* value), it notably returns between- and within-participants 98% confidence intervals (Cousineau, 2005; Cousineau & O'Brien, 2014; Morey, 2008), Cohen's d_z (Cohen, 1988) and its unbiased equivalent Hedges' g_z (Hedges, 1981; Hedges & Olkin, 1985) as well as their 98% confidence intervals (Fitts, 2020; Goulet-Pelletier & Cousineau, 2018, 2019). It also returns Cohen's d_{rm} and Hedges' g_{rm} , so that the effect sizes can easily be converted for meta-analyses.

ICA pipeline

The ICA pipeline is the alternative preprocessing pipeline. The differences with the "Original" pipeline are: Before epoching the data, a copy of the dataset is created. This copy is high-pass filtered at 2 Hz (passband edge), periods of data with no marker for more than 5000 ms are deleted and it is

then downsampled to 100 Hz. ICA weights are computed on this copy using AMICA (Palmer et al., 2008). The weights are then transferred to the original dataset. Another copy is created with a high-pass filter at 2 Hz (bandpass edge) and used for ICLabel (Pion-Tonachini et al., 2019) components classification. Components with more than 80% probability of being an eye component are flagged for rejection.

The original dataset (with ICA weights) is then epoched and the same epoch rejection as in the “Original” pipeline is performed. The eye components are then subtracted from the data and epochs with an absolute amplitude at PO7 or PO8 above 60 μV at any timepoint are additionally rejected.

Non-parametric bootstrapping

The preprocessing in this pipeline is identical to the “Original” pipeline. The differences are:

The time window of analysis is defined with a tweaked version of the collapsed localizer (Luck & Gaspelin, 2017). The collapsed localizer usually consists in averaging all participants and conditions together, and then deciding on the analysis window based on this single waveform. However, component timing in such a localizer will be more strongly affected by components with comparatively larger amplitudes (as we expect from the N2pc in the letters condition compared to the N2pc in the colors condition; see Fig. 1b) and basing the analysis window on this latency estimate would therefore bias the analyses in favor of the larger component. Thus, here we average all participants together (for a given lab) and compute the 25%-peak latencies (on- and offsets) of the difference wave for each condition (using the *latency.m* function from Liesefeld, 2018; <https://github.com/Liesefeld/latency>). We then collapse the onsets and offsets of the two N2pcs by averaging across conditions. The ipsi- and contralateral amplitudes are then extracted from this time window for each individual ERP and submitted to the same statistical test as in the “Original” pipeline. This approach allows us to obtain values that are centered on the peak, therefore better representing the “true” component independent of external factors that could impact the timing of this component (e.g., higher luminance would increase a stimulus’ salience and therefore likely result in an earlier component). However, because we search for the negative peak and create our time window based on it, this method also has the disadvantage of being biased towards finding a significant difference between contra and ipsi waves (a significant N2pc; i.e., Hypotheses 1 and 2). Therefore, we additionally run unbiased, non-parametric tests (as in e.g., Gaspelin & Luck, 2018; Liesefeld et al., 2022; Sawaki et al., 2012). Specifically, for each participant, the epoched dataset is bootstrapped (effectively assigning a random electrode laterality to each trial) and the grand average is recomputed from these bootstrapped datasets. The analysis window is derived anew at each iteration according to the above de-

scribed method. From that time window, the *negative* mean amplitude (i.e., zeroing all positive values before averaging) of the grand average ERP is extracted for each condition. We perform 10,000 iterations of this bootstrapping procedure and then compute a p value with the following equation: $p = \frac{\text{number of iterations with negative means} \leq \text{observed negative mean}}{\text{number of iterations}}$. To ensure that our p value is not the result of a lucky (or unlucky) run of the bootstrapping procedure, we repeat this procedure 1,000 times, therefore computing 1,000 p values (each from a different set of 10,000 iterations). We then keep the median p value and consider it to be the true non-parametric p value that we compare against our statistical threshold of $\alpha = 0.02$.

ICA and non-parametric bootstrapping

This pipeline combines the preprocessing of the “ICA” pipeline with the results extraction from the “Non-parametric bootstrapping” pipeline.

Known differences from the original study

While our goal is to perform a direct replication of the original study, there are some notable deviations and additional steps that we will perform and we note them here for completeness:

- The exact chromaticity values of the stimuli were not measured in the original study. Thus, we use the HSV values (converted to RGB above) of the original study (obtained through personal communication with the author and representing his best guess, because the original code was lost) and ask replicating labs to use monitors calibrated to the sRGB standard and/or measure the actual colors (xyY coordinates) produced by their setup if possible.
- During the training block, visual feedback is added in the event of an incorrect response or a timeout.
- The online reference for the EEG recording is not the right earlobe.
- The acquisition sampling rate and acquisition filters are not available in any amplifier used by the replicating labs, these settings are instead applied during offline processing. All replicating labs record the data without any filters beyond those strictly necessary for their system and with at least twice the sampling rate of the original study (i.e., 400 Hz).
- During offline preprocessing, if PO7, PO8 or a mastoid channel is flat (i.e., absolute voltage $< 4.5\text{e}-15 \mu\text{V}$) for more than 30 seconds, the participant is excluded.
- During offline preprocessing, the data is re-referenced to the average of the mastoids; this was not done in the original study but does not affect the difference between contra- and ipsilateral electrodes.
- During offline preprocessing, a bipolar VEOG channel is created by subtracting the inferior VEOG from the superior VEOG instead of subtracting the right HEOG from the superior VEOG in the original study.

- During offline preprocessing, epochs with absolute voltage from the EOGs (non-bipolar), PO7 or PO8 below 1 μV for at least 350 contiguous milliseconds are rejected.
- We do not recruit participants with a known mental disorder (recruitment criteria are not specified in the original study).
- Participants are excluded from the main analyses if they have less than 100 epochs remaining in the *forms* or *letters* condition after preprocessing.

Sample size and inclusion criteria

The most representative result are the effects of contralaterality in Study 2 (which is the replicated study) for electrode pair OL (corresponding to PO7/8 in the 10-10 system) in the time range 220 – 300 ms for form discrimination $F(1,9) = 57.10$, $p < .001$ and color discrimination $F(1,9) = 17.48$, $p = .002$; thus the smallest of these two F values (17.48) is used to compute the effect size:

$$t = \sqrt{F} = \sqrt{17.48} = 4.18$$

$$d_z = \frac{t}{\sqrt{N}} = \frac{4.18}{\sqrt{10}} = 1.32$$

Since we expect to replicate the original effect, that is, ERP amplitudes at electrodes PO7/8 are lower on the contralateral side than on the ipsilateral side, we run a one-sided paired-sample t test with the hypothesis that mean contralateral voltage < mean ipsilateral voltage (or equivalently, mean *contra minus ipsi* < 0 μV). To compute the required sample size, the package *pingouin* (version 0.5.3; Vallat, 2018) in CPython 3.10.9 was used.

As defined in the #EEGManyLabs position paper (Pavlov et al., 2021), and given that many ERP studies provide overestimated effect sizes due in part to low N s (Clayson et al., 2019), the required sample size is computed using half the effect size of the original experiment, that is a d_z of 0.66. This resulted in a required sample size of 28 participants for a one-sided paired-samples t test with an alpha of 0.02, a power of 90%. Each replicating lab commits to collect data from 28 participants. If a lab does not collect 28 participants, the data originating from that lab will not be included in the main analyses.

The recruitment criteria are:

- Older than 18 years old and older than the age of majority in the region where data is collected.
- Normal or corrected-to-normal vision
- No colorblindness
- No known mental disorder

Labs will also collect self-declared age, gender, handedness and level of education including total years and highest academic qualification of participants.

Exclusion criteria

Similar to original study:

- Epochs with an absolute VEOG above 60 μV at any time point will be excluded.
- Epochs with an absolute HEOG above 25 μV at any time point will be excluded.
- Participants with a maximal absolute residual HEOG above 2 μV will be excluded.
- Trials with an incorrect response or a timeout will be excluded.
- Trials with a target-only array are excluded from statistical analyses.

Different from original study:

- Participants with a flat (i.e., absolute voltage less than 4.5e–15 μV) mastoid electrode for more than 30 seconds are excluded.
- Epochs with an absolute voltage from the EOGs (non-bipolar), PO7 or PO8 lower than 1 μV for at least 350 contiguous milliseconds will be excluded.
- Data collection is aborted if impedances of the critical electrodes (PO7, PO8, mastoids, online reference, ground, EOGs) are not brought to a satisfactory level (see Table 1; e.g. 15 k Ω for the LMU).
- Participants with less than 100 epochs in any critical test condition (forms or colors) will be excluded.

Data sharing protocol

The raw (anonymized) data will be made available on an open access platform. Additionally, the data after marker harmonization and the epoched data will be made available at the same location. We will also share all relevant analysis scripts. Therefore each participating lab must obtain the necessary ethics approval to publicly share their data.

Confirmatory statistical analysis plan

Hypothesis 1

- Hypothesis: The mean voltage at electrode site PO7/PO8 is more negative for the electrode contralateral versus ipsilateral relative to the target’s hemifield for the form discrimination task.
- Independent variable: Electrode laterality relative to target’s hemifield (ipsilateral vs. contralateral).
- Dependent variable: Mean voltage (μV) at electrode PO7/PO8 in the defined time window.
- Time window: Time window: 220 – 300 ms for the “Original” and “ICA” pipelines. Variable (but same as H_2 and H_3) for the non-parametric bootstrapping pipelines (with or without ICA).
- Test: One-sided paired-sample t test for all pipelines; additional non-parametric test in the bootstrapping pipelines.
- Significance threshold: $p < .02$

Hypothesis 2

- Hypothesis: The mean voltage at electrode site PO7/PO8 is more negative for the electrode contralateral versus ipsilateral relative to the target’s hemifield for the color discrimination task.

- Independent variable: Electrode laterality relative to target’s hemifield (ipsilateral vs. contralateral).

- Dependent variable: Mean voltage (μV) at electrode PO7/PO8 in the defined time window.

- Time window: Time window: 220 – 300 ms for the “Original” and “ICA” pipelines. Variable (but same as H_1 and H_3) for the non-parametric bootstrapping pipelines (with or without ICA).

- Test: One-sided paired-sample t test for all pipelines; additional non-parametric test in the bootstrapping pipelines.

- Significance threshold: $p < .02$

Hypothesis 3

- Hypothesis: The mean contralateral *minus* ipsilateral voltage at electrode site PO7/PO8 is more negative for the form discrimination task than for the color discrimination task.

- Independent variable: Task (color discrimination vs. form discrimination).

- Dependent variable: Mean contralateral *minus* ipsilateral voltage (μV) at electrode PO7/PO8 in the defined time window.

- Time window: Time window: 220 – 300 ms for the “Original” and “ICA” pipelines. Variable (but same as H_1 and H_2) for the non-parametric bootstrapping pipelines (with or without ICA).

- Test: One-sided paired-sample t test for all pipelines; additional non-parametric test in the bootstrapping pipelines.

- Significance threshold: $p < .02$

Pilot data

We have collected some pilot datasets in order to test that the experimental program is functional with different setups and to develop the processing pipeline. One behavioral dataset was collected in Bremen. One EEG (and behavioral) dataset each was collected in Munich (BrainAmp DC), Kraków (BioSemi) and Essex (Neuroscan).

Data collection timeline

From the date of the in-principle-acceptance, each lab commits to collect the data within one year.

Meta-analysis

We will use a random-effects model to pool the Hedges’ g_z obtained from the “Original” pipeline of each laboratory and

their standard errors, defined as the square root of the variance computed as in Fitts (2020, Eq. 8b) with $A = (n)$ (Eq. 6b). The restricted maximum likelihood estimator (REML; Viechtbauer, 2005) will be used to estimate the heterogeneity variance τ^2 and the Knapp-Hartung adjustments (Knapp & Hartung, 2003) will be used to compute the confidence interval around the pooled effect. The meta-analysis will be computed with the R (R Core Team, 2022) package *meta* (Balduzzi et al., 2019; version 6.0.0). Replication success is defined as a statistically significant ($p < .02$) random-effects meta-analytic estimate. We will also conduct another meta-analysis with the same parameters but additionally including the original study’s effect size ($g_z = 1.21$, $SE = 0.49$).

We will report the median and distribution of the unweighted Hedges’ g_z and their 95% confidence intervals, as well as the number of datasets that successfully replicate the original effect. We will also report at least the I^2 and the prediction intervals (IntHout et al., 2016). Each Hedges’ g_z will be plotted in a forest plot. We will also report the weighted Hedges’ g_z computed like this:

$$g_z \cdot \left(\frac{1}{SE + \tau^2} / \sum \frac{1}{SE + \tau^2} \right)$$

To quantify the variation in effect sizes across samples and settings, we will further conduct a random-effects meta-analysis and establish heterogeneity estimates to determine if the amount of variability across samples exceeded the amount expected as a result of measurement error.

Contributions

Conceptualization: M.C., F.M., Y.G.P., and H.R.L.

Data curation: M.C.

Formal analysis: M.C.

Funding acquisition: F.M. and Y.G.P.

Investigation: M.C., A.M., D.A., M.Y., H.G., D.K., D.L., F.P., T.F., and H.R.L.

Methodology: M.C., F.M., Y.G.P., and H.R.L.

Project administration: F.M., Y.G.P., and H.R.L.

Software: M.C.

Supervision: F.M., Y.G.P., and H.R.L.

Validation: M.C., F.M., Y.G.P., and H.R.L.

Visualization: M.C.

Writing - original draft: M.C. and H.R.L.

Writing - review & editing: M.C., A.M., D.A., M.Y., H.G., D.K., D.L., F.P., T.F., F.M., Y.G.P., and H.R.L.

Acknowledgements

The #EEGManyLabs core team and lead lab would like to thank the original author, Martin Eimer, for generously sharing critical information to ensure the fidelity and quality of this replication attempt. We also thank Kay Robbins and Andreas Keil for their internal review of this Stage 1 Registered Report.

References

- Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, 22(4), 153–160. <https://doi.org/10/ggchpv>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10/c7g6rj>
- Burra, N., & Kerzel, D. (2013). Attentional capture during visual search is attenuated by target predictability: Evidence from the N2pc, Pd, and topographic segmentation. *Psychophysiology*, 50(5), 422–430. <https://doi.org/10/f4s98n>
- 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/ggnvxz>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10/vv3>
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42–45. <https://doi.org/10/b9z7>
- Cousineau, D., & O'Brien, F. (2014). Error bars in within-subject designs: A comment on Baguley (2012). *Behavior Research Methods*, 46(4), 1149–1151. <https://doi.org/10/f6vdsw>
- 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/bqr2f2>
- Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography & Clinical Neurophysiology*, 99(3), 225–234. <https://doi.org/10/dk5b9z>
- Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, 18(10), 526–535. <https://doi.org/10/gj6b5m>
- Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 20(8), 1423–1433. <https://doi.org/10/c69z2c>
- Fitts, D. A. (2020). Commentary on "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family": The degrees of freedom for paired samples designs. *The Quantitative Methods for Psychology*, 16(4), 281–294. <https://doi.org/10/gk3rr4>
- Gaspelin, N., & Luck, S. J. (2018). The role of inhibition in avoiding distraction by salient stimuli. *Trends in Cognitive Sciences*, 22(1), 79–92. <https://doi.org/10/gcr98d>
- Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *The Quantitative Methods for Psychology*, 14(4), 242–265. <https://doi.org/10/gkzn9m>
- Goulet-Pelletier, J.-C., & Cousineau, D. (2019). Corrigendum to "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family". *The Quantitative Methods for Psychology*, 15(1), 54–54. <https://doi.org/10/gk3pvk>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7. <https://doi.org/10/gdvhvz>
- Hedges, L. V. (1981, Summer). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107. <https://doi.org/10/dbqn45>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience*, 21(4), 760–775. <https://doi.org/10/fskhpx>
- Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, 18(4), 604–613. <https://doi.org/10/b9h54w>
- Int'Hout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7), e010247. <https://doi.org/10/ggh65k>
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36. <https://web.archive.org/web/20140812022154/http://www.perceptionweb.com/abstract.cgi?id=v070821>
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710. <https://doi.org/10/dncmwj>
- Lien, M.-C., Ruthruff, E., Goodin, Z., & Remington, R. W. (2008). Contingent attentional capture by top-down control settings: Converging evidence from event-related potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 509–530. <https://doi.org/10/fp6c6j>
- Liesefeld, H. R., Liesefeld, A. M., & Müller, H. J. (2022). Preparatory control against distraction is not feature-based. *Cerebral Cortex*, 32(11), 2398–2411. <https://doi.org/10/gmxwhc>
- Liesefeld, H. R. (2018). Estimating the timing of cognitive operations with MEG/EEG latency measures: A primer, a brief tutorial, and an implementation of various methods. *Frontiers in Neuroscience*, 12. <https://doi.org/10/gfhx3n>
- Liesefeld, H. R., Liesefeld, A. M., Töllner, T., & Müller, H. J. (2017). Attentional capture in visual search: Capture and post-capture dynamics revealed by EEG. *NeuroImage*, 156, 166–173. <https://doi.org/10/gbsjqj>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10/gdz7rc>
- Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In *The Oxford handbook of event-related potential components*. (pp. 329–360). Oxford University Press. <https://doi.org/10/jh83>
- Luck, S. J., Fan, S., & Hillyard, S. A. (1993). Attention-related modulation of sensory-evoked brain activity in a visual

- search task. *Journal of Cognitive Neuroscience*, 5(2), 188–195. <https://doi.org/10/c3cp73>
- 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/f9j228>
- Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology*, 33(1), 64–87. <https://doi.org/10/d9stvj>
- Luck, S. J., & Hillyard, S. A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. *Perception & Psychophysics*, 48(6), 603–617. <https://doi.org/10/ccnhzn>
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1000–1014. <https://doi.org/10/bj8mf5>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10/ft2dgc>
- Mitchell, M., Muftakhidinov, B., Winchen, T., Schaik, B. V., Wilms, A., Kylesower, Kensington, Jędrzejewski-Szmek, Z., Badger, T. G., & Badshah400. (2019, November 30). *Markumitchell/engauge-digitizer: Version 12.1 Directory dialogs start in saved paths* (Version v12.1). <https://doi.org/10/jnv5>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61–64. <https://doi.org/10/ggbnjg>
- Palmer, J. A., Makeig, S., Kreutz-Delgado, K., & Rao, B. D. (2008). Newton method for the ICA mixture model. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1805–1808. <https://doi.org/10/bf6sxx>
- 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/10/gjx8qt>
- 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*. <https://doi.org/10/gft89w>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10/fb9s55>
- 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/gf22bz>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. *Journal of Neuroscience*, 32(31), 10725–10736. <https://doi.org/10/f36386>
- Töllner, T., Rangelov, D., & Müller, H. J. (2012). How the speed of motor-response decisions, but not focal-attentional selection, differs as a function of task set and target prevalence. *Proceedings of the National Academy of Sciences*, 109(28). <https://doi.org/10/f35fsw>
- Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2011). Stimulus saliency modulates pre-attentive processing speed in human visual cortex (M. H. Herzog, Ed.). *PLoS ONE*, 6(1), e16276. <https://doi.org/10/bgdf2p>
- Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, 3(31), 1026. <https://doi.org/10/ggzpn5>
- van Moorselaar, D., & Slagter, H. A. (2019). Learning what is irrelevant or relevant: Expectations facilitate distractor inhibition and target facilitation through distinct neural mechanisms. *The Journal of Neuroscience*, 39(35), 6953–6967. <https://doi.org/10/gf62fg>
- Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293. <https://doi.org/10/fhd39t>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10/ggj45f>
- Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400(6747), 867–869. <https://doi.org/10/bc68bs>
- Zivony, A., Allon, A. S., Luria, R., & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: An attentional blink study. *Neuropsychologia*, 121, 153–163. <https://doi.org/10/gq4t3v>

Appendix

Table A1

Study design table

Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
Is an N2pc elicited in the form discrimination task?	The mean voltage at electrode site PO7/PO8 is more negative for the electrode contralateral versus ipsilateral relative to the target's hemifield for the form discrimination task in the time window 220 – 300 ms (for the main replication).	28 participants will be collected in each laboratory.	One-sided paired-sample <i>t</i> test for all pipelines; additional non-parametric test in the bootstrapping pipelines.	We ran a power analysis with $1 - \beta = 0.90$, $\alpha = 0.02$ and half of the replicated study's smallest effect size of interest ($d_z = 0.66$), in accordance with #EEGManyLabs recommendations.	The original finding will be deemed reliable if the meta-analytic estimate is statistically significant at $p < .02$. Conversely, the finding will be considered not replicated if the meta-analytic p value does not reach this threshold.	N/A
Is an N2pc elicited in the color discrimination task?	The mean voltage at electrode site PO7/PO8 is more negative for the electrode contralateral versus ipsilateral relative to the target's hemifield for the color discrimination task in the time window 220 – 300 ms (for the main replication).	As above.	As above.	As above.	As above.	N/A
Is the N2pc elicited in the form discrimination task larger than in the color discrimination task?	The mean contralateral <i>minus</i> ipsilateral voltage at electrode site PO7/PO8 is more negative for the form discrimination task than for the color discrimination task in the time window 220 – 300 ms (for the main replication).	As above.	As above.	As above.	As above.	N/A

Note. This table provides an overview on this replication study. Please refer to the main manuscript for details.