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

Martin Constant[1,2,3]*, Ananya Mandal[3,4], Dariusz Asanowicz[5], Bartłomiej Panek[5], Ilona Kotlewska[5], Motonori Yamaguchi[6], Helge Gillmeister[6], Dirk Kerzel[2], David Luque[7], Sara Molinero[7], Antonio Vázquez-Millán[7], Francesca Pesciarelli[8], Eleonora Borelli[8], Hanane Ramzaoui[9], Melissa Beck[9], Bertille Somon[10], Andrea Desantis[10,11,12], M. Concepción Castellanos[13], Elisa Martín-Arévalo[13], Greta Manini[13], Mariagrazia Capizzi[13], Ahu Gokce[14], Demet Özer[15,14], Efe Soyman[16], Ece Yılmaz[14], Joshua O. Eayrs[17,18], Raquel E. London[18], Tabitha Steendam[18], Christian Frings[19,20], Bernhard Pastötter[19,20], Bence Szaszkó[21,22], Pamela Baess[23], Shabnamalsadat Ayatollahi[23],Gustavo A. León Montoya[23], Nicole Wetzel[24,25], Andreas Widmann[26], Liyu Cao[27,28], Xueqi Low[27], Thiago L. Costa[29], Leonardo Chelazzi[29], Bianca Monachesi[30,29], Siri-Maria Kamp[31,20], Luisa Knopf [31,20], Roxane J. Itier[32], Johannes Meixner[33], Kerstin Jost[33], André Botes[34], Carley Braddock[34], Danqi Li[34], Alicja Nowacka[34], Marlo Quenault[34], Daniele Scanzi[34], Tamar Torrance[34], Paul M. Corballis[34], Gianvito Laera[2,35], Matthias Kliegel[2,35], Dominik Welke[36], Faisal Mushtaq[36], Yuri G. Pavlov[37], Heinrich R. Liesefeld[1,3]*

[1]University of Bremen, Bremen, Germany

[2]Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, Genève, Switzerland

[3]Graduate School of Systemic Neuroscience, Ludwig-Maximilians-Universität München, München, Germany

[4]Department of Psychology, Ludwig-Maximilians-Universität München, München, Germany

[5]Institute of Psychology, Jagiellonian University, Krakow, Poland

[6]Department of Psychology, University of Essex, Colchester, UK

[7]School of Psychology, Universidad de Málaga, Malaga, Spain

[8]University of Modena and Reggio Emilia, Modena, Italy

[9]Louisiana State University, Baton Rouge, LA, USA

[10]DTIS, ONERA, FR-13661 Salon Cedex Air, France

[11]Institut de Neurosciences de la Timone (UMR 7289), CNRS and Aix-Marseille Université, Marseille F-13005, France

[12]Integrative Neuroscience and Cognition Center (UMR 8002), CNRS and Université Paris Cité, Paris F-75006, France

[13]Mind, Brain and Behavior Research Center (CIMCYC), Department of Experimental Psychology, University of Granada, Granada, Spain

[14]Department of Psychology, Kadir Has University, Istanbul, Türkiye

[15]Department of Psychology, Bilkent University, Ankara, Türkiye

[16]Department of Psychology, Koç University, Istanbul, Türkiye

[17]Department of Psychology, Liverpool John Moores University, Liverpool, UK

[18]Department of Experimental Psychology, Ghent University, Ghent, Belgium

[19]Department of Cognitive Psychology, Trier University, Trier, Germany

[20]Institute for Cognitive and Affective Neuroscience (ICAN), Trier University, Trier, Germany

[21]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

[22]Public Mental Health Research Unit, Department of Social and Preventive Medicine, Center for Public Health, Medical University of Vienna, Austria

[23]Institute of Psychology, University of Hildesheim, Hildesheim, Germany

[24]Leibniz Institute for Neurobiology, Magdeburg, Germany

[25]Magdeburg-Stendal University of Applied Sciences, Stendal, Germany

[26]Wilhelm Wundt Institute for Psychology, Leipzig University, Leipzig, Germany

[27]Department of Psychology and Behavioural Sciences, Zhejiang University, Hangzhou, China

[28]The State Key Lab of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China

[29]Department of Neurosciences, Biomedicine and Movement, University of Verona, Verona, Italy

[30]International School for Advanced Studies - SISSA, Trieste, Italy

[31]Neurocognitive Psychology Unit, Trier University, Trier, Germany

[32]Psychology Department, University of Waterloo, Waterloo, Ontario, Canada

[33]Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[34]University of Auckland, Auckland, New Zealand

[35]Centre for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, Geneva, Switzerland

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

[37]NIHR Leeds Biomedical Research Centre, Leeds, West Yorkshire, United Kingdom

[38]University of Tuebingen, Tuebingen, Germany

*Correspondence to *martin.constant@unige.ch* or *heinrich.liesefeld@uni-bremen.de*

This version contains tracked changes (additions in red, deletions in strikethrough blue) from the first version of the Stage 2 ~~Stage 1~~ Registered Report. The clean version can be found at: https://doi.org/n6xx

The Stage 1 Registered Report (Constant et al., 2023) can be found at: https://doi.org/n6xg

The in-principle acceptance (Sherman, 2023) can be found at: https://rr.peercommunityin.org/PCIRegisteredReports/articles/rec?id=411

The OSF repository is available at: https://doi.org/n6xh

The analysis pipeline's code (Constant, 2025) is available at: https://doi.org/n3rg

The authors have no known conflicts of interest related to this manuscript.

#EEGManyLabs is funded by the DFG (PA 4005/1-1), provided to YGP and a UK Research and Innovation Biotechnology and Biological Sciences Research Council award (BB/X008428/1) to FM. Lead Replicating Senior Author (HRL) and First Author are funded by the DFG (LI 2868/3-1). AM is funded by the DFG (RTG 2175). DA, BP and IK are funded by the National Science Center of Poland (grant 2016/22/E/HS6/00139), and the Strategic Programme Excellence Initiative at Jagiellonian University (grant 221.6120.49.2023). DL is funded by the MCIN/AEI (PID2021-126767NB-I00) and the Junta de Andalucía (PROYEXCEL_00287). AD is funded by the Agence Nationale de la Recherche (ANR-18-10CE-0001). FM is supported in part by a UKRI BBSRC award (BB/X008428/1) and the

National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. LC is supported by the National Natural Science Foundation of China (32271078). RJI is funded by Natural Sciences and Engineering Research Council of Canada (NSERC). KJ is funded by the Ministry of Science, Research and Cultural Affairs of the State of Brandenburg. DK is supported by grant No. 10001E_21951 from the Swiss National Science Foundation. MY and HG received funding by the University of Essex Department of Psychology Research Promotion Fund for this study.

Contributions:

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

Data curation: M.C. and D.W.

Formal analysis: M.C.

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

Investigation: All authors

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

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

Software: M.C.

Supervision: M.C., 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 and editing: All authors

Abstract

The N2pc is widely employed as an electrophysiological marker of an attention allocation. This interpretation was largely driven by the observation of an N2pc elicited by an isolated relevant target object, which was reported as Experiment 2 in Eimer (1996). 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 strong impact on the study of selective attention.~~,~~ We revisit the idea of the N2pc being an indicator of attentional selectivity by delivering a high powered direct replication of Eimer's work through analysis of 779 datasets acquired from 22 labs across 14 countries. Our results robustly replicate the N2pc to form stimuli, but a direct replication of the ~~more influential~~ N2pc to color stimuli technically failed. We believe that this pattern not only sheds further light on the functional significance of the N2pc as an electrophysiological marker of attentional selectivity, but also highlights a methodological problem with selecting analysis windows a priori. By contrast, the consistency of observed ERP patterns across labs and analysis pipelines is stunning, and this consistency is preserved even in datasets that were rejected for (ocular) artifacts, attesting to the robustness of the ERP technique and the feasibility of large-scale multilab EEG (replication) studies.

Introduction

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, which can be valid even if it reflects some process that is a consequence of an attention allocation rather than the allocation proper. Thus, from observing an N2pc ~~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 researchers who participated in the present replication project volunteered because (a) they use or plan to use the N2pc in their work and/or (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 et al., 1993; Luck & Hillyard, 1994). 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. For example, the N2pc might reflect engagement at the location of the relevant stimulus rather than the shift of attention proper (Zivony et al., 2018). It is also possible 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).

Furthermore, 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). The target-enhancement aspect might

involve the suppression of nearby visual input if it is present (akin to Luck & Hillyard, 1994's interpretation; see Hickey et al., 2009; Wyble et al., 2020; but see also Liesefeld & Müller, 2021, Appendix D, regarding the general non-discriminability of enhancement and suppression).~~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). Target enhancement might involve the suppression of nearby visual input if it is present (akin to Luck & Hillyard, 1994's interpretation; see Hickey et al., 2009; Wyble et al., 2020; but see also Liesefeld & Müller, 2021, Appendix D, regarding the general non-discriminability of enhancement and suppression).~~

~~Furthermore, the N2pc might reflect engagement at the location of the relevant stimulus rather than the shift of attention proper (Zivony et al., 2018). It is also possible 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; Hilimire et al., 2012; 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 aimed 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 Figure 1a-b). In the following, we will refer to these conditions as "Forms" and "Colors" and to the components as "form N2pc" and "color N2pc", respectively. Thus, we aimed to replicate the two N2pcs observed in Experiment 2 of Eimer (1996; see Figure 1c-e).

Beyond these main effects of interest, a serendipitous finding is worth mentioning here: The form N2pc was larger in amplitude and temporal extent compared to the color. 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 expected to replicate a higher amplitude for an N2pc elicited by forms compared to color patches (see Figure 1e).

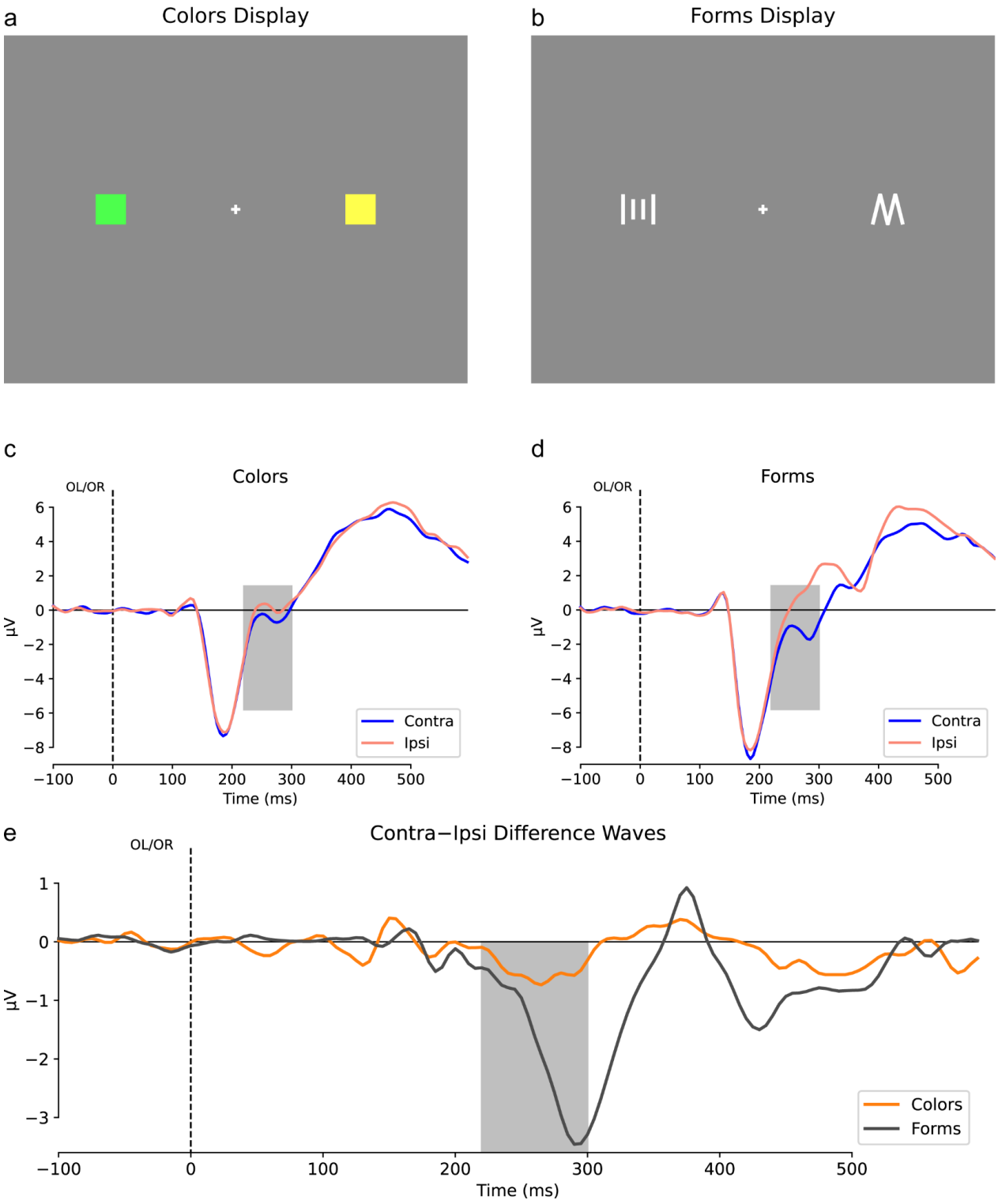


Figure 1

Displays of the experiment (a-b) and reconstructed ERPs (c-e).

(a) and (b). Search displays were recreated in OpenSesame using information from the original study's manuscript and personal communication with the author. (c) and (d). The ERPs from electrodes OL/OR (equivalent to today's PO7/PO8) were digitized from the original manuscript with Engauge (Mitchell et al., 2019), interpolated to 1000 Hz using CubicSpline interpolation with scipy v1.14.1 (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 1.9.0 (Gramfort et al., 2013), visualization was also created with MNE. The shaded area represents the original analysis time window (220 – 300 ms). Panel (e) represents the difference waves for each condition, containing the color N2pc and form N2pc. A version of this figure with inverted Y axes for panels (c), (d) and (e) is available in the OSF repository.

Methods

Transparency and openness statement

We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. The Stage 1 Registered Report (Constant et al., 2023) can be found at: https://doi.org/n6xg

The raw data (after marker harmonization and anonymization; including any complete datasets that were excluded during the analysis; Constant et al., 2025a) are available here: https://doi.org/pmg4

Additionally, the epoched data and all relevant analysis scripts (Constant et al., 2025b) are available here: https://doi.org/pmg5

Each participating lab obtained the necessary ethics approval to publicly share their data.

Stimuli, procedure & design

The experiment was developed in OpenSesame version 3.3.14 and adapted for version 4.0 (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 and the experiment are provided on https://osf.io/4ux8r/. The color values we used were obtained from personal communication with the original author and reflect his best estimate. A standard operating protocol including how to set up and run the experiment is provided in 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 distance indicated in the experimental settings is maintained], line width: 0.04 dva) was displayed against a 55% gray background for the whole experiment (i.e., it only disappeared during breaks). In half of the experimental blocks (*form discrimination* in Eimer's notation or *Forms* in ours), a letter stimulus (M or W, line width: 0.08 dva) was 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* or *Colors*), one square in a target color (blue [RGB: 30%, 30%, 100%] or green [RGB: 30%, 100%, 30%]) was presented together with a square of the same color (*target-only arrays*) or a distractor (*distractor arrays*) which was a yellow square (RGB: 100%, 100%, 30%). In each trial, the two stimuli appeared 3.3 dva to the right and left of the center of the screen for 150 ms; each stimulus subtended 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 had to indicate which target (M or W; blue or green) they saw by pressing the left or right key of their response device, independently of the target's side. The response-key assignment was counterbalanced across participants. Keypresses were stored in an asynchronous buffer. After 2150 ms this buffer was read and the first key pressed (if any) was considered to be the participant's response. Timeouts (i.e., no key pressed) were considered as errors.

As in the original study, each participant started with one condition (Forms, M vs. W, or Colors, blue vs. green; order counterbalanced) and performed 6 blocks of 66 trials of this condition before switching to the other condition with the same number of trials. There were 4 distractor-array configurations (target identity [2] × target side [2]) and there were 2 configurations for target-only arrays (target identity [2]). Each of these 6 conditions was presented an equal number of times in a block (11 times per block).

Participants were instructed not to move their eyes from the fixation cross. To train them not to move their eyes, a practice block ran until the experimenter judged from the HEOG waves that participants were holding their eyes sufficiently still. The practice block was repeated when participants started the second condition, allowing them to get accustomed with the new stimuli.

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.

The practice blocks also served as training to learn the response-key assignments and, therefore immediate feedback was provided. In particular, in the event of an incorrect response, a large gray “X” was displayed for 500 ms between two practice trials and in the event of a timeout, a gray hourglass was presented for the same duration. Correct responses did not prompt the appearance of any feedback, the fixation cross simply remained for an extra 500 ms.

EEG data acquisition

Quality assurance was undertaken by the corresponding authors for each participant lab. A video of the experimental setup as well as a pilot dataset were sent to the corresponding authors to standardize the data acquisition process as much as possible. The setup of each lab is described in Table 1.

Table 1

Overview of the EEG set-up and recording details at each replicating lab

Participating university	***N* collected** ***N* in Original** ***N* in ICA**	**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	34 28 26	BrainProducts BrainAmp DC 1000 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 15 kΩ	REF: FCz GND: Fpz	HP: 0.016 Hz 1st order 6dB/octave LP: 250 Hz 5th order Butterworth 30dB/octave	Windows XP BrainVision recorder (v1.20.0601)	50 Hz	VIEWPixx/3D (1920×1080, 120Hz, scanning backlight)	Windows 10	Course credits or 10 €/h
Jagiellonian University (Krakow)	37 26 26	BioSemi ActiveTwo Mk2 1024 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 208 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v7)	50 Hz	Samsung SyncMaster 2243 (1920×1080, 60 Hz)	Windows 10	50 zł/h
University of Essex	39 28 28	Compumedics Neuroscan SynAmps RT 1000 Hz	Ag/AgCl EasyCap (26 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 15 kΩ	REF: M1 GND: AFz	HP: 0.05 Hz 6dB/octave LP: 100 Hz 2nd order Butterworth	Windows 10 Curry 8	50 Hz	Dell S2419HGF (1920×1080, 120 Hz)	Windows 10	Course credits or 8 £/h
Université de Genève (Kerzel)	35 27 24	BrainProducts actiCHamp 1000 Hz	Ag/AgCl ActiCap Snap (26 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 10 kΩ	REF: FCz GND: AFz	HP: DC LP: 280 Hz	Windows 10 BrainVision Recorder (v1.25.0204)	50 Hz	VIEWPixx Lite (1920×1200, 100 Hz, normal backlight)	Windows 10	Course credits
Universidad de Málaga	38 28 26	BrainProducts BrainAmp DC 500 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG +	REF: FCz GND: Fpz	HP: 0.016 Hz 1st order 6dB/octave LP: 1000	Windows 10 BrainVision Recorder (v1.24.0101)	50 Hz	Lenovo G24qe-20 (2560×1440, 60	Windows 10	10 €/h

			2 mastoids) 15 kΩ		Hz 5th order Butterworth 30dB/octave			Hz)		
University of Modena and Reggio Emilia (UNIMORE)	30 20 20	BrainProducts actiCHamp Plus 1000 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 20 kΩ	REF: FCz GND: Fpz	HP: DC LP: 280 Hz	Windows 10 BrainVision Recorder (v1.25.0101)	50 Hz	Philips 107B (1024×768, 60 Hz, 230×306 mm)	Windows 10	Course credits
Louisiana State University (LSU)	42 25 22	BioSemi ActiveTwo Mk2 512 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 104 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v7.2)	60 Hz	BenQ XL2420-b (1920×1080, 60 Hz)	Windows 10	Course credits
ONERA The French Aerospace Lab	38 23 23	BrainProducts ActiCHamp 1000 Hz	Ag/AgCl ActiCap Snap (58 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 10 kΩ	REF: FCz GND: Fpz	HP: DC LP: 280 Hz	Windows 7 BrainVision Recorder (v1.25.0202)	50 Hz	LG Flatron 915 FTPlus (1024×768, 60 Hz)	Windows 7	15 €/h
University of Granada (NCC_UGR)	38 27 27	BrainProducts ActiCHamp 500 Hz	Ag/AgCl ActiCap Snap (26 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 10 kΩ	REF: Cz GND: Fpz	HP: DC LP: 140 Hz	Windows 10 BrainVision Recorder (v1.25.0201)	50 Hz	BenQ BL2405 (1920×1080, 60 Hz)	Window 10	10 €/h
Kadir Has University (KHas)	29 16 15	BrainProducts ActiCHamp 1000 Hz	Ag/AgCl ActiCap Snap (26 scalp + 2 HEOG + 2 VEOG + 2 mastoids) 10 kΩ	REF: Cz GND: Fpz	HP: DC LP: 280 Hz	Windows 10 BrainVision Recorder (v1.22.0001)	50 Hz	MSI G241V (1920×1080, 75 Hz)	Windows 10	Course credits or 75 TL/h
Ghent University	29 10 9	BioSemi ActiveTwo 512 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 104 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v8.0)	50 Hz	BenQ XL2411z (1920×1080, 60 Hz)	Windows 10	Course credits or 12 €/h
Trier University (Pastötter,	28 12 12	BrainProducts BrainAmp DC 500 Hz	Ag/AgCl (57 scalp + 2 HEOG	REF: FCz GND: AFz	HP: 0.016 Hz 1st order	Windows 7 Pro BrainVision Recorder (v1.20.0801)	50 Hz	EIZO S1911 (1280×10	Windows 7	Course credits or 15 €/h

Frings; TrierCogPsy)			+ 2 VEOG + 2 mastoids) 20 kΩ		6dB/octave LP: 1000 Hz 5th order Butterworth 30dB/octave			24, 60 Hz)		
University of Vienna	36 24 24	BioSemi ActiveTwo 512 Hz	Ag/AgCl (128 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 104 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v9.02)	50 Hz	Sony GDM-F500R (1600×1200, 75 Hz)	Windows 10	Course credits
University of Hildesheim	32 28 28	BioSemi ActiveTwo 512 Hz	Ag/AgCl custom-made (32 scalp + 2 HEOG + 2 VEOG + 2 mastoids + nose + right earlobe)	REF: CMS GND: DRL	HP: DC LP: 104 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v9.02)	50 Hz	Dell G2422HS (1920×1080, 165 Hz)	Windows 10	Course credits or 12 €/h
Leibniz Institute for Neurobiology, Magdeburg	33 25 24	BrainProducts ActiCHamp 500 Hz	Ag/AgCl ActiCap Snap (56 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 20 kΩ	REF: Nose tip GND: Fpz	HP: DC LP: 140 Hz	Windows 10 BrainVision Recorder (v1.25.0202)	50 Hz	VIEWPixx/EEG (1920×1080, 120 Hz, scanning backlight)	Ubuntu Linux 22.04	Course credits or 10 €/h
Zhejiang University (ZJU)	35 27 27	BioSemi ActiveTwo 1024 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 208 Hz 5th order CIC filter	Windows 11 BioSemi ActiView (v8.09-Beta)	50Hz	HP X24ih (1920×1080, 60 Hz)	Windows 10	RMB 50/h
Verona University	29 27 26	BrainProducts ActiCHamp Plus 1000 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 20 kΩ	REF: Fz GND: Fpz	LP: DC HP: 280 Hz	Windows 10 BrainVision Recorder (v1.24.0001)	50 Hz	AOC M2470SWH (1920×1080, 60 Hz)	Windows 10	10 €/h
Trier University (Kamp)	39 28 28	NeurOne Tesla VP00430 500Hz	Ag/AgCl (14 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 10 kΩ	REF: FCz GND: AFz	HP: 0.16 Hz LP: 125 Hz	Windows 7 NeurOne (v1.4.1.64)	50 Hz	LG 24MB37PM (1920×1080, 60 Hz)	Windows 7	Course credits or 12 €/h
University of Waterloo (ItierLab)	62 42 41	BioSemi ActiveTwo 512 Hz	Ag/AgCl custom-made (66 scalp + 2 HEOG + 2 VEOG +	REF: CMS GND: DRL	HP: DC LP: 104 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v7.07)	60 Hz	ViewSonic G90fB (1280×1024, 85 Hz)	Windows 10	Course credits

			2 mastoids)							
Brandenburg Medical School Theodor Fontane, Neuruppin	29 27 27	BrainProducts actiCHamp 1000 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 25 kΩ	REF: FCz GND: Fpz	HP: DC LP: 280 Hz	Windows 10 BrainVision Recorder (v1.23.0004)	50 Hz	Alienware AW2521HF (1920×1080, 240 Hz)	Windows 10	Course credits or 10 €/h
University of Auckland	34 21 20	BrainProducts actiCHamp Plus 1000 Hz	Ag/AgCl ActiCap Snap (59 scalp + 2 HEOG + 1 VEOG + 2 mastoids) 10 kΩ	REF: FCz GND: Fpz	HP: DC LP: 280 Hz	Windows 10 BrainVision Recorder (v1.23.0003)	50 Hz	LG 24MK600M (1920×1080, 60 Hz)	Windows 10	Course credits or 20 NZD/h
Université de Genève (Kliegel)	34 21 16	BioSemi ActiveTwo 2048 Hz	Ag/AgCl (64 scalp + 2 HEOG + 2 VEOG + 2 mastoids)	REF: CMS GND: DRL	HP: DC LP: 417 Hz 5th order CIC filter	Windows 10 BioSemi ActiView (v9.02)	50 Hz	BenQ XL2420Z (1920×1080, 60 Hz)	Windows 10	Course credits

Note. BioSemi amplifiers do not allow measuring the impedances, therefore there is no impedance threshold for labs using these amplifiers.

EEG offline processing

The EEG data were preprocessed with two slightly different pipelines and results were 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 were used to cross-validate the results with more modern processing techniques. The analysis code (Constant, 2025) is available at https://doi.org/n3rg.

Original pipeline

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

EEG data were imported from the original recording format to EEGLAB (2024.0; Delorme & Makeig, 2004). After import, the markers were cleaned and harmonized to a

common scheme, and markers reflecting the reaction time were added from information contained in the behavioral file.

At this point, for the purpose of flatline (channel blocking) detection only, a copy of the dataset was created and high-passed filtered at 1 Hz (bandpass edge) with "pop_eegfiltnew(EEG, 'locutoff', 1, 'usefftfilt', 1)" (Widmann et al., 2015) and with periods of data where no marker was sent for more than 5000 ms removed. If a mastoid electrode or PO7 or PO8 was flat (absolute voltage < 4.5e-15μV) for more than 30 seconds in this copied dataset, the participant was excluded and further processing was not performed.

Next, the electrode layout in the original data set was harmonized (i.e., referenced to the BESA template) and data were re-referenced to the average of the mastoids. Data were 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, 'usefftfilt', 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, 'usefftfilt', 0)". Finally, data were 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 were created (baseline correction: -100ms – 0 ms). Only epochs for distractor arrays where the participant's response was correct were created. A bipolar horizontal EOG channel was created by subtracting the right HEOG from the left HEOG and a bipolar vertical EOG channel was 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 a-voltage from the EOGs (non-bipolar), PO7 or PO8 below ±1 µV for at least 350 contiguous milliseconds were rejected. Epochs were also rejected if the amplitude of the bipolar VEOG was larger than ±60 µV or if the amplitude of the bipolar HEOG was larger than ±25 µV at any timepoint in the epoch. The data were then averaged with ERPLAB (12.00; Lopez-Calderon & Luck, 2014). The left and right EOG- and EEG-electrodes were then converted to contralateral or ipsilateral electrodes and contralateral *minus* ipsilateral difference waves were created. At this point, if the maximal voltage of the HEOG difference wave, in the ERP calculated across all conditions, exceeded ±2 µV at any time point, the participant was 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 were then extracted and statistically analyzed with paired-samples *t* tests (see Confirmatory analysis plan).

The paired-samples *t* test was performed with a custom implementation in MATLAB 2024a 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.

In addition to these frequentist *t* tests, we performed directed Bayes Factor (*BF*) *t* tests with the BayesFactor (version 0.9.12-4.7; Morey & Rouder, 2024) R package (version 4.4.1; R Core Team, 2024), which is equivalent to running them with JASP (0.19.1; JASP Team, 2024; Love et al., 2019) with default settings for the prior (half Cauchy distribution with a mode of 0 and a width of $\frac{\sqrt{2}}{2}$). A *BF* in favor of the null ≥ 3 (i.e., $BF_{10} \leq 1/3$) or a *BF* in favor of the alternative ≥ 6 was considered as sufficient evidence.

We also report the robustness check performed with the BayesFactor R package (i.e., changing the width of the Cauchy distribution to 1.0 and to 1.4). In the event that frequentist statistics and *BF*s results diverge, we draw our conclusions from the frequentist statistics (following the general approach of the #EEGManyLabs project; Pavlov et al., 2021).

ICA pipeline

The ICA pipeline is the alternative preprocessing pipeline and conforms more closely to the approach taken in many current N2pc studies. The differences to the "Original" pipeline are:

Before epoching the data, a copy of the dataset was created. This copy was high-pass filtered at 2 Hz (passband edge), periods of data with no marker for more than 5000 ms were deleted and it was then downsampled to 100 Hz. ICA weights were computed on this copy using AMICA (1.7; Palmer et al., 2008). The weights were then transferred to the original dataset.

Another copy-was created with a high-pass filter at 2 Hz (bandpass edge, one-pass, zero-phase, non-causal FIR filter, Hamming-windowed sinc, filter order 331) and used for ICLabel (1.6.0; Pion-Tonachini et al., 2019) components classification. Components with more than 80% probability of being an eye component were flagged for rejection.

379 The original dataset (with ICA weights) was then epoched and the same participant and epoch
380 rejection as in the "Original" pipeline were performed. The eye components were then subtracted
381 from the data and epochs with an amplitude at PO7 or PO8 exceeding ±60 µV at any timepoint
382 were additionally rejected (thus yielding a higher number of rejected trials and - consequently -
383 rejected participants compared to the original pipeline).

384 ***Collapsed localizer pipeline***

385 The preprocessing in this pipeline was identical to the "Original" pipeline, but instead of
386 using a fixed time window, this pipeline uses an objective approach to adapt the time windows to
387 the empirical data (Luck & Gaspelin, 2017). The differences are:

388 The time window of analysis was defined with a tweaked version of the collapsed
389 localizer (Luck & Gaspelin, 2017). The collapsed localizer usually consists of averaging all
390 participants and conditions together, and then deciding on the analysis window based on this
391 single waveform. However, component timing in such a localizer is more strongly affected by
392 components with comparatively larger amplitudes (as we expected from the form N2pc
393 compared to the color N2pc; see Figure 1e) and basing the analysis window on this latency
394 estimate would therefore bias the analyses in favor of the larger component. Thus, we estimated
395 latencies separately for each condition (based on the grand average in each lab) and collapsed
396 afterwards across conditions. On- and offsets were quantified as 25% of the maximal amplitude
397 of the strongest negative component in the difference wave (in a 100 – 350 ms search window
398 using the *latency.m* function from Liesefeld, 2018; https://github.com/Liesefeld/latency). We
399 then collapsed the onsets and offsets of the two N2pcs by averaging across conditions. The ipsi-
400 and contralateral amplitudes were then extracted from this time window for each individual ERP
401 and submitted to the same statistical test as in the "Original" pipeline.

We expected that this approach would allow us to obtain values that are centered on the N2pc 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 in the contra-ipsi difference wave 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 ran 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 was bootstrapped (effectively assigning a random electrode laterality to each trial) and the grand average was recomputed from these bootstrapped datasets. The analysis window was derived anew at each iteration according to the above described method. From that time window, the *negative* mean amplitude (i.e., zeroing all positive values before averaging) of the grand average ERP was extracted for each condition. We performed 10,000 iterations of this bootstrapping procedure and then computed a *p* value with the following equation:

$$p = \frac{number\ of\ iterations\ with\ negative\ means \leq observed\ negative\ mean}{number\ of\ iterations}$$

To ensure that our *p* value was not the result of a lucky (or unlucky) run of the bootstrapping procedure, we repeated this procedure 1,000 times, therefore computing 1,000 *p* values (each from a different set of 10,000 iterations). We then kept the median *p* value (henceforth: p_{boot}) and considered it to be the true non-parametric *p* value that we compared against our statistical threshold of $\alpha = .02$.

ICA and collapsed localizer pipeline

This pipeline combined the preprocessing of the “ICA” pipeline with the results extraction from the “Collapsed localizer” pipeline.

Known differences from the original study

While our goal was to perform a direct replication of the original study, there were some notable deviations and additional steps that we performed 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 asked replicating labs to use monitors calibrated to the sRGB colorspace and/or measure the actual colors (xyY coordinates) produced by their setup if possible.
- During the training block, visual feedback was added in the event of an incorrect response or a timeout.
- The acquisition sampling rate and acquisition filters used in the original study were not available in any amplifier used by the replicating labs; comparable settings were instead applied during offline processing. All replicating labs recorded 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 was flat (i.e., absolute voltage < 4.5e-15 μV) for more than 30 seconds, the participant was excluded.
- The online reference for the EEG recording was not the right earlobe for any lab. During offline preprocessing, the data were 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 was 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 voltage from the EOGs (non-bipolar), PO7 or PO8 below ±1 μV for at least 350 contiguous milliseconds were rejected.
- We did not recruit participants with a known mental disorder (recruitment criteria are not specified in the original study).
- Participants were excluded from the main analyses if they had less than 100 epochs remaining in Forms or Colors after preprocessing.

Sample size and inclusion criteria

The most influential results of Eimer (1996) are the effects of contralaterality in Experiment 2 (which is the replicated study) for electrode pair OL/OR (corresponding to PO7/PO8 in the 10-10 system) in the time range 220 – 300 ms. Experiment 2 is, in a sense, more influential than Experiment 1, because with only one non-target item, it provides a stronger test of the main hypothesis that the N2pc is related to target processing rather than the suppression of

surrounding non-targets. The spatiotemporal extent of this effect is most influential as it corresponds most closely to the typical analysis window of the N2pc in subsequent studies.

We aimed to replicate three effects which are the form and color N2pcs as well as the difference in amplitude between the two. In the original study, these are reflected by the main effects of contralaterality, $F(1, 9) = 57.10$, $p < .001$ and $F(1, 9) = 17.48$, $p = .002$ and the interaction of task with contralaterality, $F(1, 9) = 37.49$, $p < .001$, respectively. Thus the smallest of these F values (17.48) was 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 expected to replicate the original effect, that is, ERP amplitudes at electrodes PO7/PO8 are more negative ~~lower~~ on the contralateral side than on the ipsilateral side, we ran a one-sided paired-samples *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 was 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 committed to collect data from 28 participants. If a lab did not collect 28 participants, the data originating from that lab were not included in the main analyses. We note that one lab included in Stage 1 was unable

to collect any data and is therefore removed from Table 1 in this Stage 2 Report. The recruitment criteria were:

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

Labs also collected age, gender, handedness and level of education including total years and highest academic qualification of participants. These data, including the ones pertaining to recruitment criteria were self-declared by the participants.

Exclusion criteria

Similar to the original study:

- Epochs with a VEOG exceeding ±60 µV at any time point were excluded.
- Epochs with a HEOG exceeding ±25 µV at any time point were excluded.
- Participants with a maximal residual HEOG exceeding ±2 µV were excluded.
- Trials with an incorrect response or a timeout were excluded.
- Trials with a target-only array were excluded from statistical analyses.

Different from the original study:

- Participants with a flat (i.e., absolute voltage less than 4.5e-15 µV) mastoid electrode for more than 30 seconds were excluded.

- Epochs with a voltage from the EOGs (non-bipolar), PO7 or PO8 lower than ±1 µV for at least 350 contiguous milliseconds were excluded.
- Data collection was aborted if impedances of the critical electrodes (PO7, PO8, mastoids, online reference, ground, EOGs) were not brought to a satisfactory level (see Table 1; e.g. 15 kΩ for the LMU). Since BioSemi amplifiers do not allow the measure of impedances, this was not an exclusion criterion for labs which used them.
- Participants with less than 100 epochs in any critical test condition (Forms or Colors) were excluded.

~~**Data sharing protocol**~~

~~The raw (anonymized) data (including any complete datasets that were excluded during the analysis) are available here: https://cloud.fak11.lmu.de/index.php/s/pm6wtZQTjPTpFHo. Additionally, the data after marker harmonization and the epoched data are available at the same location. We also share all relevant analysis scripts. Each participating lab obtained 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 **Forms** (i.e., there is a form N2pc).
- 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: 220 – 300 ms for the "Original" and "ICA" pipelines. Variable (but same as H_2 and H_3) for the collapsed localizer pipelines (with or without ICA).
- Test: One-sided paired-samples *t* test for all pipelines (frequentist and Bayes Factor); additional non-parametric test in the collapsed localizer pipelines.
- Significance threshold: $p < .02$; $BF_{10} \geq 6$ or $BF_{10} \leq 1/3$ is considered as substantial evidence for the alternative or null hypothesis, respectively.

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 **Colors** (i.e., there is a color N2pc).
- 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: 220 – 300 ms for the "Original" and "ICA" pipelines. Variable (but same as H_2 and H_3) for the collapsed localizer pipelines (with or without ICA).
- Test: One-sided paired-samples *t* test for all pipelines (frequentist and Bayes Factor); additional non-parametric test in the collapsed localizer pipelines.
- Significance threshold: $p < .02$; $BF_{10} \geq 6$ or $BF_{10} \leq 1/3$ is considered as substantial evidence for the alternative or null hypothesis, respectively.

Hypothesis 3:

- Hypothesis: The mean contralateral *minus* ipsilateral voltage at electrode site PO7/PO8 is more negative for Forms than Colors (i.e., the form N2pc is larger in amplitude than the color N2pc).
- Independent variable: Task/Condition (Colors vs. Forms).
- Dependent variable: Mean voltage (µV) at electrode PO7/PO8 in the defined time window.
- Time window: 220 – 300 ms for the "Original" and "ICA" pipelines. Variable (but same as H_2 and H_3) for the collapsed localizer pipelines (with or without ICA).
- Test: One-sided paired-samples *t* test for all pipelines (frequentist and Bayes Factor); additional non-parametric test in the collapsed localizer pipelines.
- Significance threshold: $p < .02$; $BF_{10} \geq 6$ or $BF_{10} \leq 1/3$ is considered as substantial evidence for the alternative or null hypothesis, respectively.

Pilot data

We collected pilot data to test that the experimental program was 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).

Meta-analysis

For each pipeline, we used a random-effects model to pool the Hedges' g_z obtained from each lab 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) was used to estimate the heterogeneity variance τ^2 and the Knapp-Hartung adjustment (Knapp & Hartung, 2003) was used to compute the confidence interval around the pooled effect. The meta-analysis was computed with the R (version 4.4.1; R Core Team, 2024) package *meta* (Balduzzi et al., 2019; version 7.0.0). Replication success was defined as a statistically significant ($p < .02$) random-effects meta-analytic estimate. For the "Original" pipeline, we also conducted another meta-analysis with the same parameters but additionally including the original study's effect size (Colors: $g_z = -1.21$, $SE = 0.49$, Forms: $g_z = -2.18$, $SE = 0.73$, Difference: $g_z = -1.77$, $SE = 0.62$).

We report the median and each lab's unweighted Hedges' g_z and their 98% confidence intervals, as well as the number of datasets that successfully replicate the original effect. We also report at least the I^2 and the prediction intervals (IntHout et al., 2016). Each Hedges' g_z is plotted in a forest plot. We also report the weighted Hedges' g_z computed with the following formula:

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

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

Results

In the following, we first report and interpret the results from the planned pipelines. A more "deliberate" and common – though less principled – approach to the analysis of these data is provided further below.

Participants and exclusion

Overall, 22 labs contributed at least 28 participants (before exclusion by the "Original" pipeline). Some labs tested extra participants to try to reach 28 participants after exclusion by the pipeline. This resulted in data from 779 participants, of which 538 (69.1%) remained after exclusion in the "Original" pipeline. In that pipeline and the "Collapsed localizer" pipeline (which shares the same preprocessing), the minimum number of participants per lab after exclusion was 10 and the maximum was 42 (M = 24.5). In the ICA pipeline, we expected to reject more participants since we added one exclusion criterion for trials. This supplementary rejection criterion led to 19 more participants being excluded, for a remaining number of 519 participants (66.6%). For the non-excluded participants in the Original pipeline, there was an average of 29.54% rejected trials for Forms and 33.29% for Colors. In the ICA pipeline, these were 29.78% and 33.73% respectively.

Original pipeline

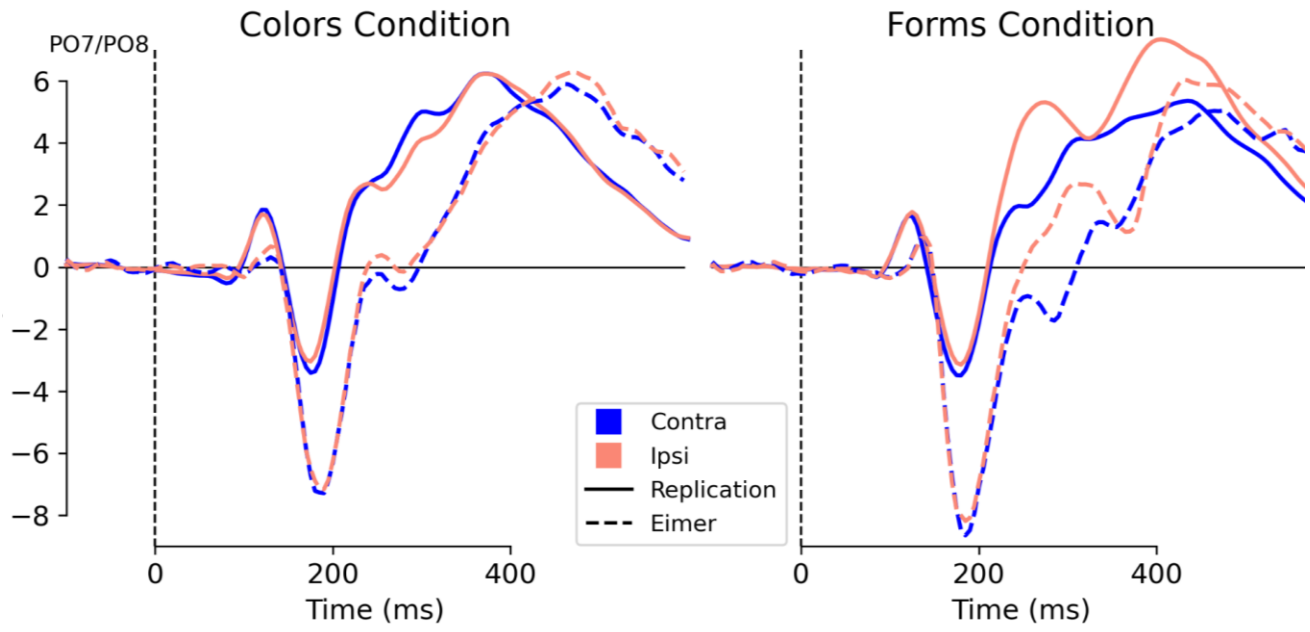


Figure 2. *Contra- and ipsi-lateral waveforms for both conditions*.

Note. Data were first averaged across trials, then across participants, and finally across labs. In our replication, the N1 latency was 175 ms for Colors and 180 ms for Forms. The P1 latencies in our replication were 120 ms and 125 ms for Colors and Forms respectively. Based on the reconstructed data, the N1 latencies in the original study were 190 ms for Colors and 185 ms for Forms. For P1, they were at 130 ms and 140 ms respectively.

Against our firm convictions, the color N2pc did not replicate in any lab (see Table 2, Figure 2 and Figure 3). To our surprise, the *BF* evidence for the null hypothesis exceeded our threshold of 1/3 for all 22 labs. Moreover, the effect was in the opposite direction than expected, with the amplitude being greater on the contralateral side compared to the ipsilateral side. The median g_z was 0.58. As expected, the form N2pc replicated in all labs. The *BF* evidence for the alternative hypothesis was above our threshold of 6 for all 22 labs. The median g_z was -1.48. As expected, the Difference between form and color N2pc replicated in all labs. That is, in all labs, the form N2pc was more negative than the color N2pc. The *BF* evidence for the alternative hypothesis was above our threshold of 6 for all labs. The median g_z was -1.62.

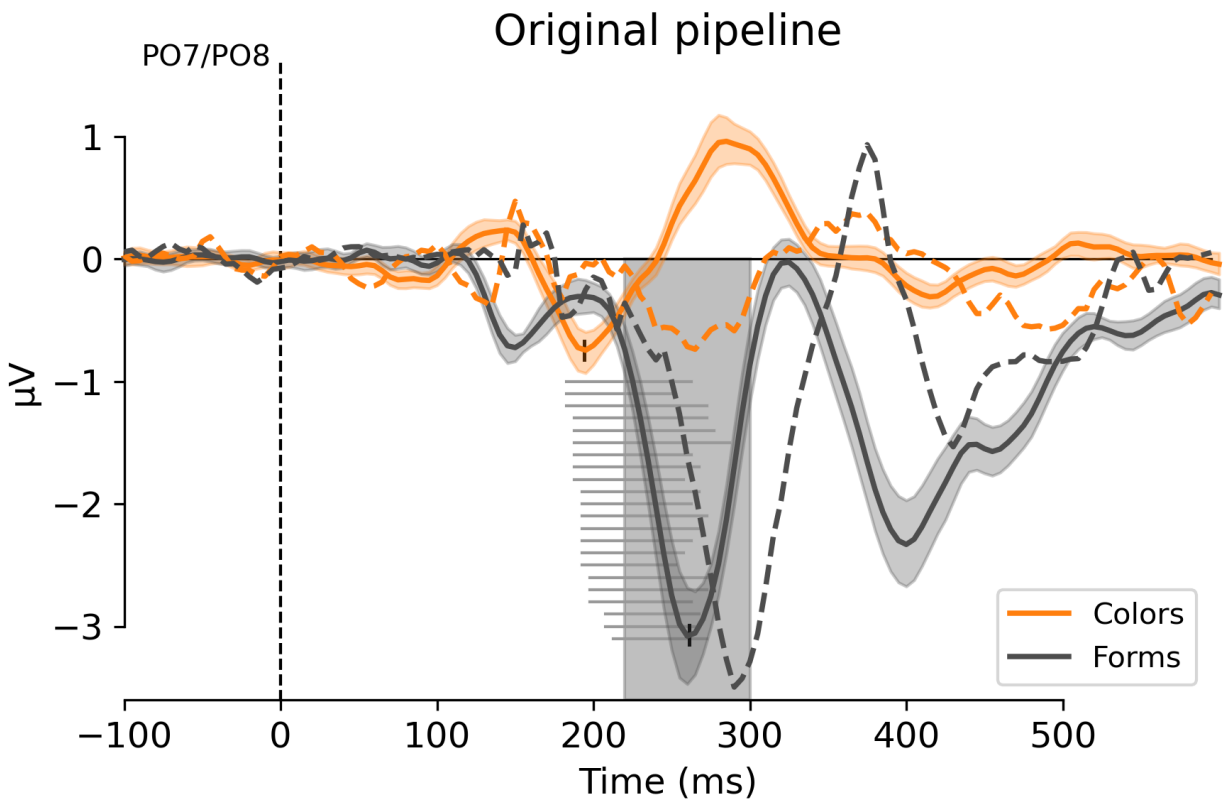


Figure 3. *Grand average difference waves for the "Original" preprocessing pipelines.*

Note. The plain lines with the shaded area (98% confidence interval) reflect the average difference wave of each lab's grand average. The dashed lines represent the reconstructed difference wave from the original study (as in Figure 1, panel b). The analysis window for the "Collapsed localizer" pipeline varies across labs and is represented by the thin horizontal gray lines (1 line per lab). The small black vertical lines represent what we deem to be the peaks of the color and form N2pcs. Each lab's individual ERP with both time windows displayed (common and individual) is also available in the OSF repository.

Table 2. *Results from the "Original" pipeline*

Lab	***t***	***df***	***p***	g_z **[98% *CI*]**	BF_{-0} **[wide, ultrawide]**
Colors					

Auckland	3.14	20	.997	0.66 [0.15, 1.38]	0.07 [0.05, 0.03]
Essex	2.77	27	.995	0.51 [0.07, 1.07]	0.06 [0.04, 0.03]
GenevaKerzel	8.25	26	> .999	1.54 [1.00, 2.44]	0.04 [0.03, 0.02]
GenevaKliegel	3.98	18	> .999	0.87 [0.33, 1.73]	0.06 [0.05, 0.03]
Gent	3.66	9	.997	1.06 [0.31, 2.64]	0.10 [0.07, 0.05]
Hildesheim	3.88	27	> .999	0.71 [0.27, 1.32]	0.05 [0.04, 0.03]
ItierLab	2.30	41	.987	0.35 [-0.01, 0.76]	0.05 [0.04, 0.03]
KHas	-0.27	15	.396	-0.06 [-0.73, 0.57]	0.32 [0.24, 0.17]
Krakow	3.73	25	> .999	0.71 [0.25, 1.35]	0.05 [0.04, 0.03]
LSU	1.62	24	.940	0.31 [-0.16, 0.87]	0.09 [0.06, 0.05]
Magdeburg	3.79	24	> .999	0.73 [0.26, 1.40]	0.05 [0.04, 0.03]
Malaga	4.68	27	> .999	0.86 [0.41, 1.51]	0.05 [0.03, 0.02]
Munich	2.11	27	.978	0.39 [-0.05, 0.92]	0.07 [0.05, 0.04]
NCC_UGR	2.17	26	.980	0.41 [-0.04, 0.95]	0.07 [0.05, 0.04]
Neuruppin	2.65	26	.993	0.50 [0.05, 1.06]	0.06 [0.05, 0.03]
ONERA	3.48	22	.999	0.70 [0.21, 1.39]	0.06 [0.04, 0.03]
TrierCogPsy	3.36	11	.997	0.90 [0.23, 2.14]	0.09 [0.07, 0.05]
TrierKamp	2.02	27	.973	0.37 [-0.07, 0.90]	0.07 [0.05, 0.04]
UNIMORE	4.94	19	> .999	1.06 [0.51, 1.96]	0.06 [0.04, 0.03]

UniversityofVienna	2.00	23	.971	0.40 [-0.08, 0.98]	0.08 [0.06, 0.04]
Verona	2.20	26	.982	0.41 [-0.04, 0.96]	0.07 [0.05, 0.04]
ZJU	0.24	26	.594	0.05 [-0.43, 0.53]	0.17 [0.12, 0.09]
Forms					
Auckland	-7.31	20	< .001	-1.53 [-2.60, -0.93]	7.32e+04 [8.65e+04, 9.31e+04]
Essex	-6.41	27	< .001	-1.18 [-1.93, -0.69]	4.84e+04 [5.31e+04, 5.25e+04]
GenevaKerzel	-7.47	26	< .001	-1.40 [-2.24, -0.87]	4.69e+05 [5.43e+05, 5.67e+05]
GenevaKliegel	-8.35	18	< .001	-1.83 [-3.14, -1.15]	2.27e+05 [2.80e+05, 3.18e+05]
Gent	-4.63	9	.001	-1.34 [-3.16, -0.56]	68.69 [76.43, 77.98]
Hildesheim	-9.85	27	< .001	-1.81 [-2.78, -1.23]	1.08e+08 [1.31e+08, 1.48e+08]
ItierLab	-6.72	41	< .001	-1.02 [-1.56, -0.63]	6.73e+05 [7.08e+05, 6.68e+05]
KHas	-7.45	15	< .001	-1.77 [-3.22, -1.05]	1.84e+04 [2.25e+04, 2.52e+04]
Krakow	-10.46	25	< .001	-1.99 [-3.09, -1.36]	1.56e+08 [1.96e+08, 2.29e+08]
LSU	-5.90	24	< .001	-1.14 [-1.94, -0.64]	9138.32 [9912.81, 9683.92]
Magdeburg	-10.47	24	< .001	-2.03 [-3.18, -1.38]	1.03e+08 [1.30e+08, 1.52e+08]
Malaga	-7.72	27	< .001	-1.42 [-2.25, -0.90]	1.04e+06 [1.21e+06, 1.28e+06]

Munich	-9.02	27	< .001	-1.66 [-2.57, -1.10]	1.84e+07 [2.23e+07, 2.47e+07]
NCC_UGR	-7.25	26	< .001	-1.35 [-2.18, -0.84]	2.82e+05 [3.24e+05, 3.35e+05]
Neuruppin	-5.57	26	< .001	-1.04 [-1.76, -0.56]	5623.46 [5924.06, 5617.08]
ONERA	-7.03	22	< .001	-1.42 [-2.37, -0.85]	7.17e+04 [8.31e+04, 8.71e+04]
TrierCogPsy	-7.00	11	< .001	-1.88 [-3.82, -1.04]	2093.52 [2569.76, 2918.61]
TrierKamp	-8.41	27	< .001	-1.54 [-2.42, -1.01]	4.92e+06 [5.85e+06, 6.33e+06]
UNIMORE	-8.89	19	< .001	-1.91 [-3.20, -1.22]	8.09e+05 [1.01e+06, 1.16e+06]
UniversityofVienna	-9.13	23	< .001	-1.80 [-2.89, -1.18]	5.88e+06 [7.25e+06, 8.21e+06]
Verona	-5.63	26	< .001	-1.05 [-1.78, -0.57]	6393.78 [6757.47, 6427.44]
ZJU	-5.84	26	< .001	-1.09 [-1.83, -0.61]	1.05e+04 [1.12e+04, 1.08e+04]
				Difference	
Auckland	-7.67	20	< .001	-1.61 [-2.72, -0.99]	1.44e+05 [1.73e+05, 1.89e+05]
Essex	-6.76	27	< .001	-1.24 [-2.01, -0.75]	1.10e+05 [1.23e+05, 1.24e+05]
GenevaKerzel	-9.06	26	< .001	-1.69 [-2.65, -1.12]	1.47e+07 [1.79e+07, 1.99e+07]
GenevaKliegel	-9.34	18	< .001	-2.05 [-3.47, -1.32]	1.06e+06 [1.34e+06, 1.57e+06]
Gent	-5.65	9	< .001	-1.63 [-3.71, -0.79]	215.84 [254.26, 276.01]
Hildesheim	-11.72	27	< .001	-2.15 [-3.26, -1.51]	3.93e+09 [5.02e+09, 5.99e+09]
ItierLab	-5.97	41	< .001	-0.90 [-1.42, -0.52]	6.66e+04 [6.73e+04, 6.11e+04]
KHas	-5.76	15	< .001	-1.37 [-2.59, -0.72]	1359.64 [1545.41, 1596.24]

Krakow	-14.63	25	< .001	-2.78 [-4.22, -1.99]	1.49e+11 [1.97e+11, 2.49e+11]
LSU	-6.24	24	< .001	-1.21 [-2.03, -0.69]	1.98e+04 [2.19e+04, 2.18e+04]
Magdeburg	-10.65	24	< .001	-2.06 [-3.22, -1.41]	1.41e+08 [1.78e+08, 2.10e+08]
Malaga	-9.40	27	< .001	-1.73 [-2.67, -1.16]	4.14e+07 [5.06e+07, 5.67e+07]
Munich	-8.69	27	< .001	-1.60 [-2.49, -1.05]	9.04e+06 [1.09e+07, 1.19e+07]
NCC_UGR	-7.63	26	< .001	-1.43 [-2.28, -0.90]	6.74e+05 [7.84e+05, 8.26e+05]
Neuruppin	-5.82	26	< .001	-1.09 [-1.83, -0.60]	1.02e+04 [1.09e+04, 1.05e+04]
ONERA	-8.06	22	< .001	-1.62 [-2.66, -1.03]	5.54e+05 [6.66e+05, 7.31e+05]
TrierCogPsy	-7.15	11	< .001	-1.92 [-3.90, -1.07]	2480.50 [3059.40, 3495.76]
TrierKamp	-7.75	27	< .001	-1.42 [-2.26, -0.90]	1.11e+06 [1.29e+06, 1.36e+06]
UNIMORE	-9.04	19	< .001	-1.94 [-3.25, -1.25]	1.04e+06 [1.30e+06, 1.50e+06]
UniversityofVienna	-8.90	23	< .001	-1.76 [-2.82, -1.15]	3.78e+06 [4.63e+06, 5.21e+06]
Verona	-7.18	26	< .001	-1.34 [-2.17, -0.83]	2.42e+05 [2.77e+05, 2.86e+05]
ZJU	-5.17	26	< .001	-0.97 [-1.66, -0.49]	2117.75 [2173.84, 2011.46]

Note. Since we expected a negativity, directed *t* tests and BF_{-0} (quantifying the evidence for the directed, negative, hypothesis) are reported here and in the following. Note that only negative *t* values could be significant.

Meta-analysis

The random-effects meta-analytic estimate for Colors was $t(21) = 7.86$, $p > .999$ (see Figure 4), after adding the original effect size to the meta-analysis, the estimate was $t(22) = 6.20$, $p > .999$, therefore this effect was not replicated. For Forms, the estimate was $t(21) = –19.99$, $p <$

.001 (see Figure 5), after adding the original effect size to the meta-analysis, the estimate was $t(22) = -20.13$, $p < .001$, therefore this effect was replicated. For the Difference between conditions, the estimate was $t(21) = -16.81$, $p < .001$ (see Figure 6), after adding the original effect size to the meta-analysis, the estimate was $t(22) = -17.34$, $p < .001$, therefore this effect was replicated as well.

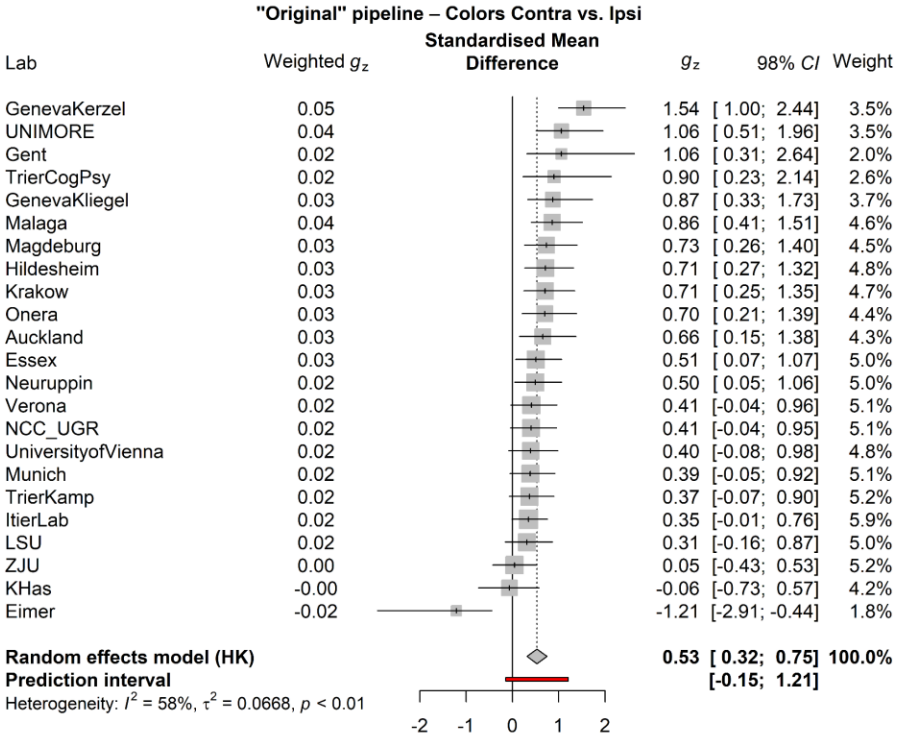


Figure 4. *Forest plot of the meta-analysis for Colors in the "Original" pipeline.*

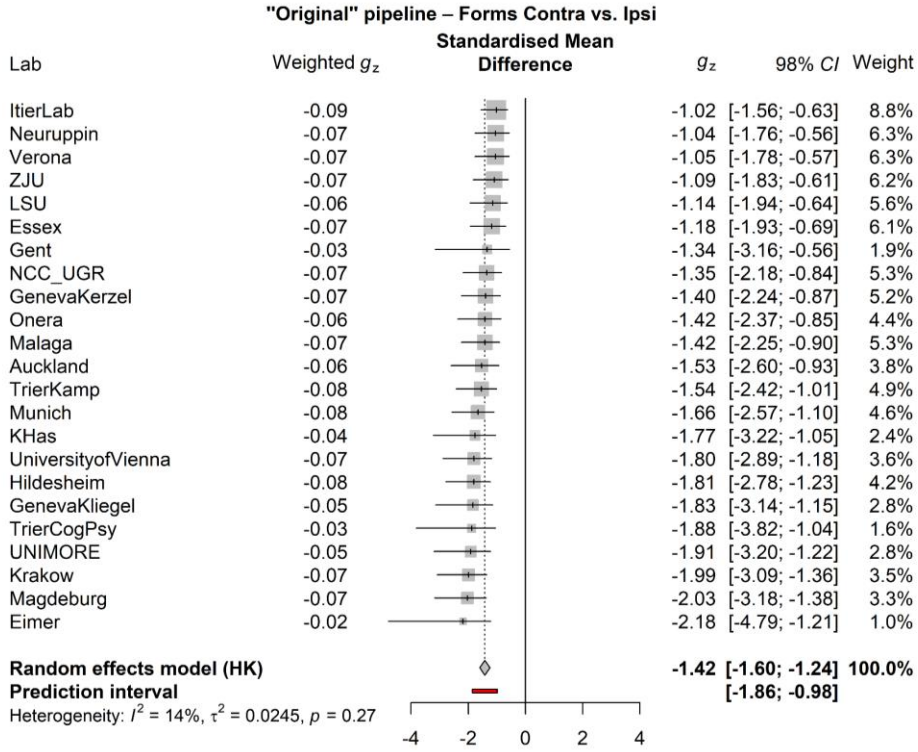


Figure 5. *Forest plot of the meta-analysis for Forms in the "Original" pipeline.*

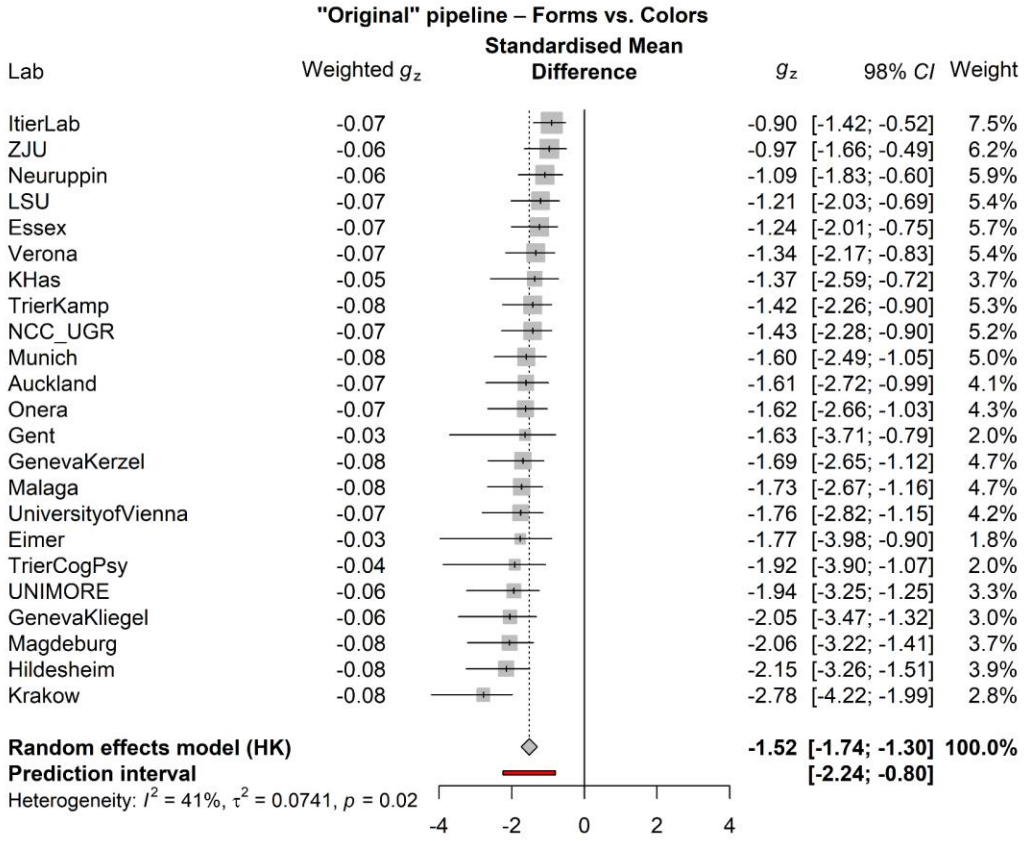


Figure 6. *Forest plot of the meta-analysis for Difference in the "Original" pipeline.*

ICA pipeline

The color N2pc did not replicate in any lab (see Table 3 and Figure 7). Again, the *BF* evidence for the null hypothesis exceeded our threshold of 1/3 for all labs. The median g_z was 0.55. The form N2pc replicated in all labs. The *BF* evidence for the alternative hypothesis was above our threshold of 6 for all labs. The median g_z was -1.48. The Difference between form and color N2pc replicated in all labs. The *BF* evidence for the alternative hypothesis was above our threshold of 6 for all labs. The median g_z was -1.54.

Table 3. *Results from the "ICA" pipeline.*

Lab	*t*	*df*	*p*	g_z [98% *CI*]	BF_{-0} [wide, ultrawide]

			Colors		
Auckland	2.79	19	.994	0.60 [0.08, 1.32]	0.07 [0.05, 0.04]
Essex	2.80	27	.995	0.51 [0.07, 1.07]	0.06 [0.04, 0.03]
GenevaKerzel	7.58	23	> .999	1.50 [0.93, 2.45]	0.04 [0.03, 0.02]
GenevaKliegel	2.64	15	.991	0.63 [0.05, 1.48]	0.09 [0.06, 0.04]
Gent	2.45	8	.980	0.74 [-0.04, 2.20]	0.13 [0.09, 0.07]
Hildesheim	4.03	27	> .999	0.74 [0.29, 1.36]	0.05 [0.03, 0.02]
ItierLab	2.14	40	.981	0.33 [-0.04, 0.74]	0.06 [0.04, 0.03]
KHas	-0.09	14	.464	-0.02 [-0.71, 0.65]	0.28 [0.21, 0.15]
Krakow	3.55	25	.999	0.67 [0.22, 1.30]	0.06 [0.04, 0.03]
LSU	2.35	21	.986	0.48 [-0.01, 1.13]	0.08 [0.05, 0.04]
Magdeburg	3.57	23	.999	0.70 [0.23, 1.38]	0.06 [0.04, 0.03]
Malaga	5.00	25	> .999	0.95 [0.47, 1.66]	0.05 [0.03, 0.02]
Munich	2.82	25	.995	0.54 [0.08, 1.13]	0.06 [0.04, 0.03]
NCC_UGR	2.30	26	.985	0.43 [-0.02, 0.98]	0.07 [0.05, 0.04]
Neuruppin	2.56	26	.992	0.48 [0.03, 1.04]	0.07 [0.05, 0.03]
ONERA	3.54	22	.999	0.71 [0.22, 1.41]	0.06 [0.04, 0.03]
TrierCogPsy	3.26	11	.996	0.87 [0.20, 2.09]	0.10 [0.07, 0.05]
TrierKamp	1.91	27	.967	0.35 [-0.09, 0.87]	0.08 [0.05, 0.04]

UNIMORE	4.98	19	> .999	1.07 [0.52, 1.97]	0.06 [0.04, 0.03]
UniversityofVienna	2.18	23	.980	0.43 [-0.05, 1.02]	0.08 [0.05, 0.04]
Verona	3.67	25	.999	0.70 [0.24, 1.33]	0.05 [0.04, 0.03]
ZJU	0.18	26	.571	0.03 [-0.44, 0.51]	0.18 [0.13, 0.09]
Forms					
Auckland	-7.15	19	< .001	-1.53 [-2.64, -0.92]	4.11e+04 [4.86e+04, 5.23e+04]
Essex	-6.44	27	< .001	-1.18 [-1.93, -0.70]	5.13e+04 [5.65e+04, 5.59e+04]
GenevaKerzel	-9.61	23	< .001	-1.90 [-3.02, -1.26]	1.43e+07 [1.78e+07, 2.04e+07]
GenevaKliegel	-7.99	15	< .001	-1.90 [-3.43, -1.15]	4.01e+04 [4.97e+04, 5.68e+04]
Gent	-4.10	8	.002	-1.24 [-3.15, -0.43]	30.02 [32.47, 32.28]
Hildesheim	-9.87	27	< .001	-1.81 [-2.79, -1.23]	1.10e+08 [1.37e+08, 1.55e+08]
ItierLab	-6.92	40	< .001	-1.06 [-1.62, -0.66]	1.10e+06 [1.18e+06, 1.13e+06]
KHas	-8.47	14	< .001	-2.07 [-3.79, -1.26]	4.85e+04 [6.11e+04, 7.16e+04]
Krakow	-11.12	25	< .001	-2.12 [-3.27, -1.46]	5.18e+08 [6.60e+08, 7.83e+08]
LSU	-5.52	21	< .001	-1.13 [-2.00, -0.60]	2595.80 [2801.94, 2728.33]
Magdeburg	-9.88	23	< .001	-1.95 [-3.10, -1.30]	2.30e+07 [2.88e+07, 3.34e+07]
Malaga	-7.40	25	< .001	-1.41 [-2.28, -0.87]	3.19e+05 [3.70e+05, 3.88e+05]
Munich	-8.52	25	< .001	-1.62 [-2.57, -1.05]	3.49e+06 [4.21e+06, 4.62e+06]

NCC_UGR	-7.50	26	< .001	-1.40 [-2.25, -0.88]	5.05e+05 [5.85e+05, 6.12e+05]
Neuruppin	-5.74	26	< .001	-1.07 [-1.80, -0.59]	8300.60 [8829.92, 8451.94]
ONERA	-7.13	22	< .001	-1.44 [-2.40, -0.87]	8.85e+04 [1.03e+05, 1.09e+05]
TrierCogPsy	-7.12	11	< .001	-1.91 [-3.88, -1.07]	2402.17 [2960.15, 3378.53]
TrierKamp	-8.57	27	< .001	-1.57 [-2.46, -1.03]	6.93e+06 [8.29e+06, 9.02e+06]
UNIMORE	-7.88	19	< .001	-1.69 [-2.88, -1.05]	1.50e+05 [1.81e+05, 2.01e+05]
UniversityofVienna	-8.88	23	< .001	-1.75 [-2.82, -1.14]	3.68e+06 [4.50e+06, 5.06e+06]
Verona	-7.11	25	< .001	-1.35 [-2.20, -0.83]	1.67e+05 [1.91e+05, 1.98e+05]
ZJU	-5.79	26	< .001	-1.08 [-1.82, -0.60]	9381.11 [1.00e+04, 9609.00]

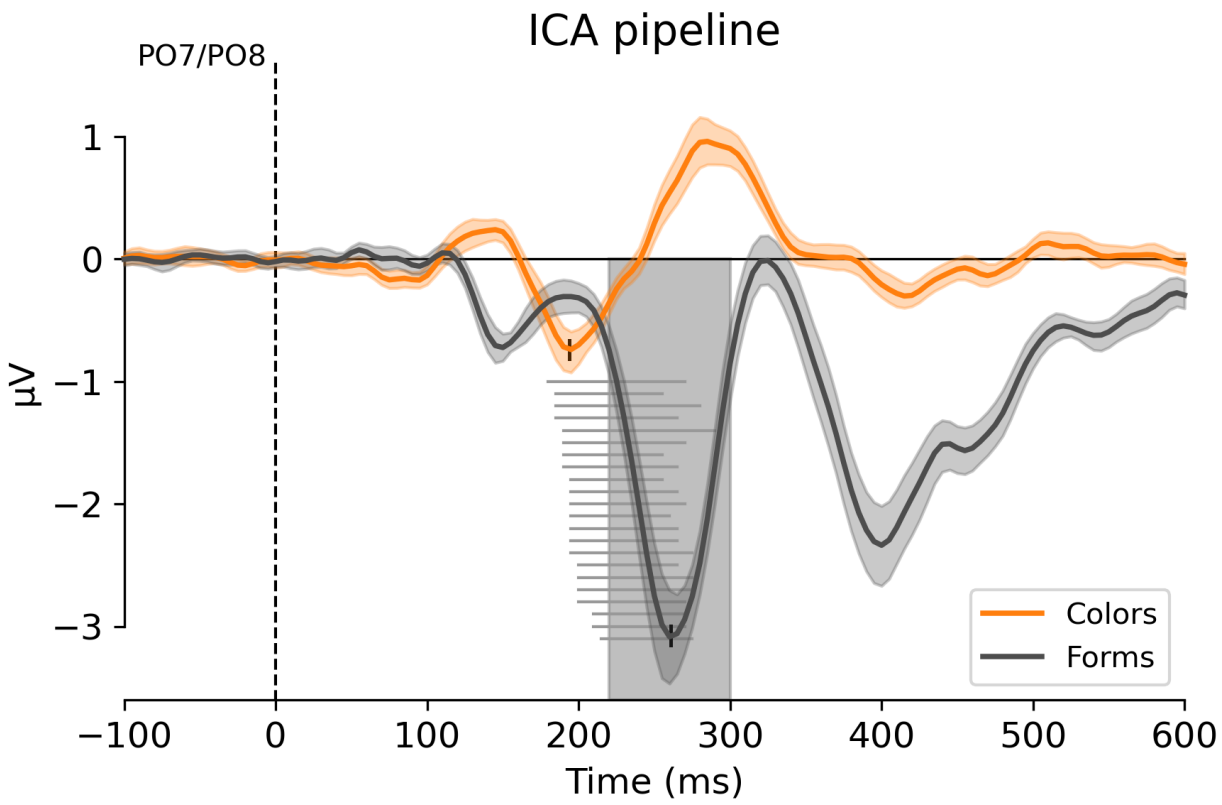
Difference

Auckland	-7.42	19	< .001	-1.59 [-2.73, -0.96]	6.67e+04 [7.96e+04, 8.66e+04]
Essex	-6.75	27	< .001	-1.24 [-2.01, -0.75]	1.09e+05 [1.22e+05, 1.22e+05]
GenevaKerzel	-10.80	23	< .001	-2.13 [-3.36, -1.45]	1.15e+08 [1.46e+08, 1.74e+08]
GenevaKliegel	-7.80	15	< .001	-1.85 [-3.35, -1.11]	3.03e+04 [3.73e+04, 4.24e+04]
Gent	-4.55	8	.001	-1.37 [-3.41, -0.55]	48.98 [54.66, 56.10]
Hildesheim	-11.80	27	< .001	-2.17 [-3.28, -1.52]	4.63e+09 [5.93e+09, 7.08e+09]
ItierLab	-6.19	40	< .001	-0.95 [-1.48, -0.56]	1.22e+05 [1.25e+05, 1.15e+05]
KHas	-6.80	14	< .001	-1.66 [-3.13, -0.94]	5047.67 [6048.16, 6643.46]

Krakow	-13.32	25	< .001	-2.53 [-3.86, -1.79]	2.03e+10 [2.67e+10, 3.30e+10]
LSU	-6.47	21	< .001	-1.33 [-2.28, -0.77]	1.85e+04 [2.10e+04, 2.15e+04]
Magdeburg	-10.03	23	< .001	-1.98 [-3.14, -1.33]	3.02e+07 [3.80e+07, 4.42e+07]
Malaga	-8.99	25	< .001	-1.71 [-2.70, -1.13]	9.02e+06 [1.10e+07, 1.23e+07]
Munich	-9.08	25	< .001	-1.73 [-2.72, -1.14]	1.08e+07 [1.32e+07, 1.48e+07]
NCC_UGR	-8.01	26	< .001	-1.50 [-2.38, -0.96]	1.55e+06 [1.83e+06, 1.95e+06]
Neuruppin	-5.94	26	< .001	-1.11 [-1.85, -0.62]	1.34e+04 [1.44e+04, 1.39e+04]
ONERA	-8.10	22	< .001	-1.63 [-2.67, -1.03]	5.97e+05 [7.18e+05, 7.89e+05]
TrierCogPsy	-7.25	11	< .001	-1.95 [-3.95, -1.10]	2791.97 [3454.77, 3963.86]
TrierKamp	-7.78	27	< .001	-1.43 [-2.26, -0.91]	1.20e+06 [1.39e+06, 1.47e+06]
UNIMORE	-8.25	19	< .001	-1.77 [-2.99, -1.11]	2.81e+05 [3.45e+05, 3.88e+05]
UniversityofVienna	-8.81	23	< .001	-1.74 [-2.80, -1.13]	3.21e+06 [3.92e+06, 4.40e+06]
Verona	-7.77	25	< .001	-1.48 [-2.37, -0.93]	7.10e+05 [8.34e+05, 8.88e+05]
ZJU	-4.97	26	< .001	-0.93 [-1.62, -0.46]	1320.41 [1337.40, 1222.53]

Meta-analysis

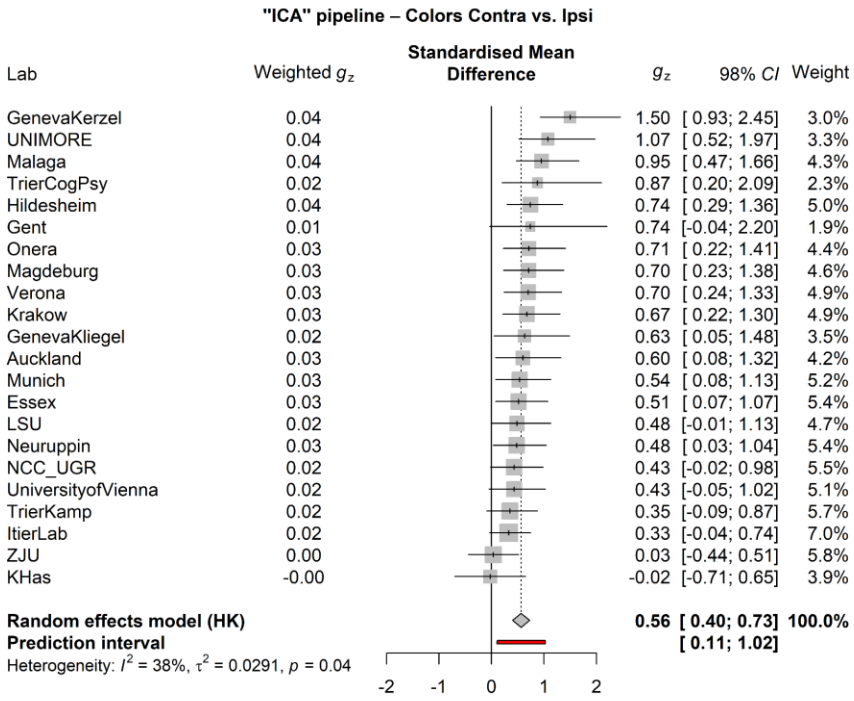
The random-effects meta-analytic estimate for Colors was $t(21) = 7.71$, $p > .999$ (see Figure 8), therefore this effect was not replicated. For Forms, the estimate was $t(21) = –20.49$, $p < .001$ (see Figure 9), therefore this effect was replicated. For the difference between conditions, the estimate was $t(21) = –17.86$, $p < .001$ (see Figure 10) and therefore this effect was also replicated.



677

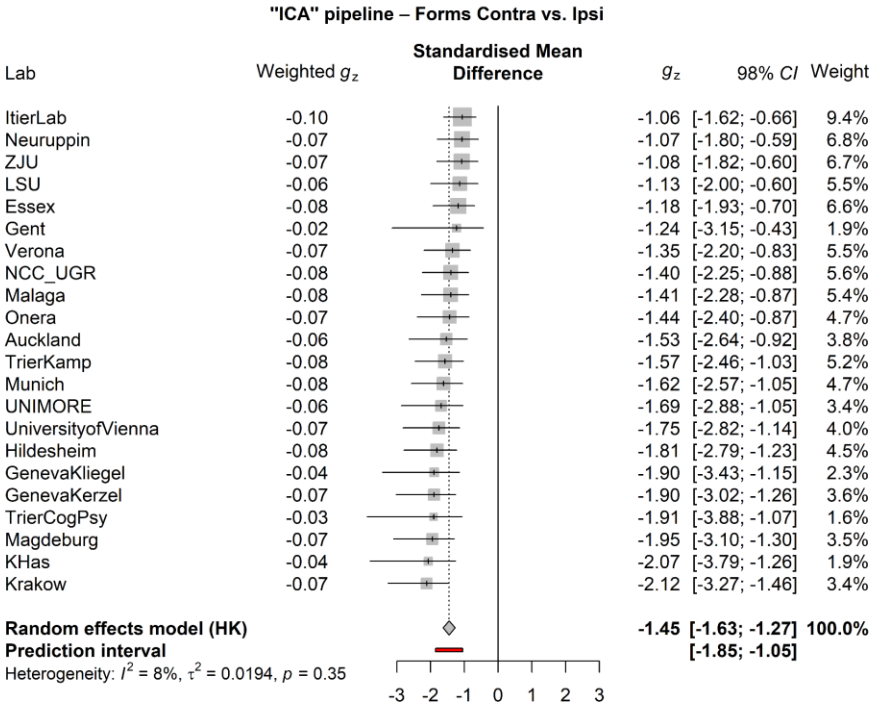
678 Figure 7. *Grand average difference waves for the "ICA" preprocessing pipelines.*

679 *Note.* The plain lines with the shaded area (98% confidence interval) reflect the average difference wave of each
680 lab's grand average. Note that these difference waves are shared with the "ICA & collapsed localizer" pipeline. The
681 analysis window for that pipeline varies across labs and is represented by the thin horizontal gray lines (1 line per
682 lab). The small black vertical lines represent what we deem to be the peaks of the color and form N2pcs.



683

684 Figure 8. *Forest plot of the meta-analysis for Colors in the "ICA" pipeline.*



685

686 Figure 9. *Forest plot of the meta-analysis for Forms in the "ICA" pipeline.*

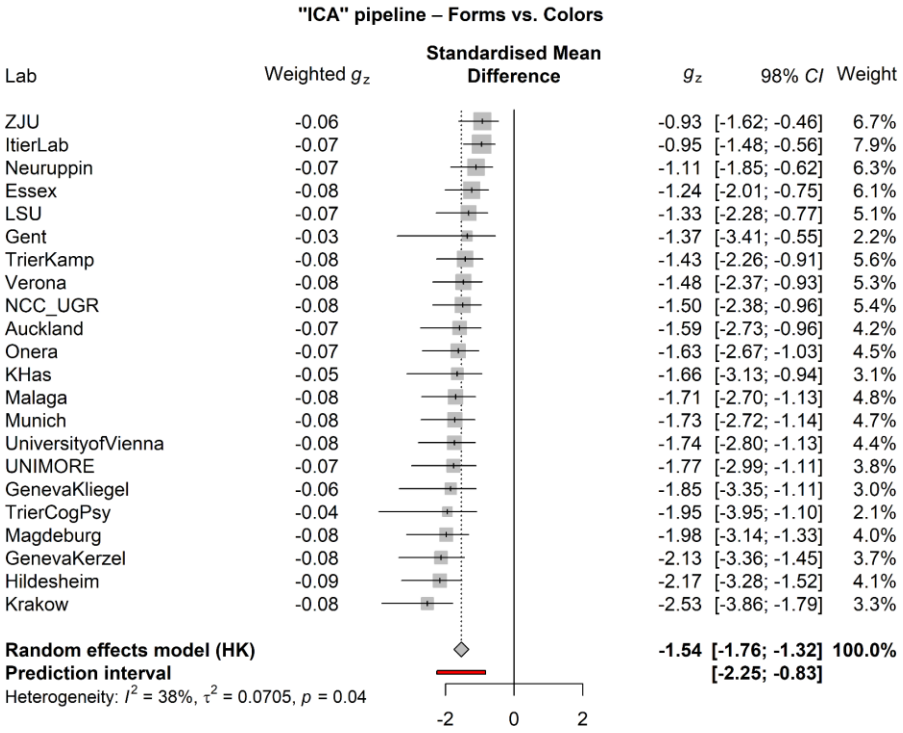


Figure 10. *Forest plot of the meta-analysis for Difference in the "ICA" pipeline.*

Collapsed localizer pipeline

We searched for the 25% onset and offset amplitude latency between 100 and 350 ms for each condition, and averaged the two resulting onsets. The time windows are available in Table 4. Note, that we had originally used a search window between 100 and 450 ms, but for four teams, the function considered the late negative peak as the form N2pc (because it was larger in amplitude than the negative peak in the typical N2pc time window), which led to largely delayed estimates. This also applies to the ICA & Collapsed localizer pipeline.

The color N2pc replicated in 16 labs out of 22 (see Table 4). The median g_z was -0.16. The form N2pc replicated in all labs. The median g_z was -1.14. The Difference between form and color N2pc replicated in all labs. The median g_z was -0.92.

699

700 Table 4. *Results from the collapsed localizer pipeline.*

Lab	Time window	*t*	*df*	p_{boot}	g_z [98% *CI*]	BF_{-0} [wide, ultrawide]
			Colors			
Auckland	185 – 265 ms	-1.35	20	.009	-0.28 [-0.89, 0.23]	0.90 [0.70, 0.52]
Essex	200 – 275 ms	-0.14	27	.090	-0.03 [-0.50, 0.44]	0.22 [0.16, 0.12]
GenevaKerzel	195 – 255 ms	1.53	26	.196	0.29 [-0.17, 0.81]	0.09 [0.06, 0.05]
GenevaKliegel	195 – 265 ms	-0.35	18	.016	-0.08 [-0.68, 0.50]	0.32 [0.23, 0.17]
Gent	190 – 280 ms	0.84	9	.285	0.24 [-0.57, 1.25]	1.03 [0.84, 0.66]
Hildesheim	210 – 270 ms	-0.16	27	.073	-0.03 [-0.50, 0.43]	0.23 [0.17, 0.12]
ItierLab	215 – 275 ms	-0.63	41	.019	-0.10 [-0.48, 0.27]	0.03 [0.02, 0.02]
KHas	200 – 275 ms	-3.02	15	.006	-0.72 [-1.61, -0.13]	12.35 [11.40, 9.70]
Krakow	195 – 260 ms	0.54	25	.085	0.10 [-0.37, 0.60]	0.14 [0.10, 0.07]
LSU	190 – 275 ms	-1.26	24	.025	-0.24 [-0.78, 0.23]	0.06 [0.04, 0.03]
Magdeburg	190 – 260 ms	-0.95	24	.015	-0.18 [-0.71, 0.30]	0.51 [0.39, 0.28]
Malaga	200 – 265 ms	0.62	27	.001	0.11 [-0.34, 0.60]	0.13 [0.10, 0.07]
Munich	195 – 270 ms	-2.44	27	< .001	-0.45 [-0.99, -0.01]	4.80 [3.95, 3.07]
NCC_UGR	195 – 275 ms	-1.29	26	< .001	-0.24 [-0.75, 0.21]	0.76 [0.58, 0.43]
Neuruppin	195 – 265 ms	-0.39	26	.004	-0.07 [-0.56, 0.39]	0.28 [0.21, 0.15]

ONERA	195 – 270 ms	-1.90	22	< .001	-0.38 [-0.98, 0.10]	1.95 [1.56, 1.20]
TrierCogPsy	190 – 270 ms	0.18	11	.046	0.05 [-0.72, 0.85]	0.25 [0.19, 0.14]
TrierKamp	210 – 270 ms	-0.83	27	.013	-0.15 [-0.64, 0.30]	0.43 [0.32, 0.23]
UNIMORE	190 – 265 ms	-1.34	19	< .001	-0.29 [-0.92, 0.24]	0.90 [0.70, 0.53]
UniversityofVienna	185 – 255 ms	-2.69	23	< .001	-0.53 [-1.15, -0.06]	7.62 [6.50, 5.19]
Verona	185 – 275 ms	-0.96	26	.002	-0.18 [-0.68, 0.28]	0.51 [0.38, 0.28]
ZJU	190 – 290 ms	-0.94	26	< .001	-0.18 [-0.68, 0.28]	0.25 [0.18, 0.13]
			Forms			
Auckland	185 – 265 ms	-4.30	20	< .001	-0.90 [-1.71, -0.38]	183.86 [183.58, 166.25]
Essex	200 – 275 ms	-5.03	27	< .001	-0.92 [-1.59, -0.46]	1652.63 [1671.92, 1526.03]
GenevaKerzel	195 – 255 ms	-3.72	26	< .001	-0.70 [-1.31, -0.24]	70.45 [64.83, 54.69]
GenevaKliegel	195 – 265 ms	-6.82	18	< .001	-1.50 [-2.64, -0.88]	1.76e+04 [2.06e+04, 2.20e+04]
Gent	190 – 280 ms	-4.62	9	< .001	-1.34 [-3.15, -0.55]	72.78 [81.25, 83.17]
Hildesheim	210 – 270 ms	-6.02	27	< .001	-1.11 [-1.83, -0.63]	1.85e+04 [1.99e+04, 1.92e+04]
ItierLab	215 – 275 ms	-6.11	41	< .001	-0.93 [-1.44, -0.54]	19.77 [16.42, 12.81]
KHas	200 – 275 ms	-7.44	15	< .001	-1.77 [-3.22, -1.04]	1.82e+04 [2.22e+04, 2.49e+04]
Krakow	195 – 260 ms	-10.61	25	< .001	-2.02 [-3.13, -1.38]	2.04e+08 [2.58e+08, 3.02e+08]
LSU	190 – 275 ms	-5.25	24	< .001	-1.02 [-1.77, -0.52]	45.62 [41.81, 35.18]

Magdeburg	190 – 260 ms	-8.39	24	< .001	-1.62 [-2.60, -1.05]	1.94e+06 [2.33e+06, 2.56e+06]
Malaga	200 – 265 ms	-7.38	27	< .001	-1.36 [-2.17, -0.85]	4.76e+05 [5.47e+05, 5.66e+05]
Munich	195 – 270 ms	-7.18	27	< .001	-1.32 [-2.12, -0.81]	2.97e+05 [3.38e+05, 3.47e+05]
NCC_UGR	195 – 275 ms	-6.10	26	< .001	-1.14 [-1.90, -0.65]	1.98e+04 [2.15e+04, 2.10e+04]
Neuruppin	195 – 265 ms	-4.49	26	< .001	-0.84 [-1.50, -0.38]	421.39 [412.31, 365.62]
ONERA	195 – 270 ms	-5.71	22	< .001	-1.15 [-2.00, -0.62]	4513.15 [4898.08, 4792.56]
TrierCogPsy	190 – 270 ms	-5.80	11	< .001	-1.56 [-3.25, -0.79]	491.95 [575.60, 617.43]
TrierKamp	210 – 270 ms	-5.33	27	< .001	-0.98 [-1.66, -0.51]	3439.78 [3549.75, 3299.41]
UNIMORE	190 – 265 ms	-10.46	19	< .001	-2.24 [-3.71, -1.49]	8.90e+06 [1.14e+07, 1.37e+07]
UniversityofVienna	185 – 255 ms	-7.04	23	< .001	-1.39 [-2.30, -0.84]	9.23e+04 [1.06e+05, 1.11e+05]
Verona	185 – 275 ms	-5.87	26	< .001	-1.10 [-1.84, -0.61]	1.13e+04 [1.21e+04, 1.17e+04]
ZJU	190 – 290 ms	-4.91	26	< .001	-0.92 [-1.60, -0.45]	3596.74 [3745.31, 3512.20]

Difference

Auckland	185 – 265 ms	-3.74	20	< .001	-0.78 [-1.55, -0.27]	57.38 [54.66, 47.55]
Essex	200 – 275 ms	-4.30	27	< .001	-0.79 [-1.42, -0.34]	281.29 [269.81, 235.20]
GenevaKerzel	195 – 255 ms	-3.67	26	< .001	-0.68 [-1.30, -0.23]	62.10 [56.89, 47.81]
GenevaKliegel	195 – 265 ms	-5.98	18	< .001	-1.31 [-2.36, -0.72]	3808.50 [4294.84, 4386.04]
Gent	190 – 280 ms	-5.71	9	< .001	-1.65 [-3.75, -0.81]	46.97 [51.13, 51.00]

Hildesheim	210 – 270 ms	-5.29	27	< .001	-0.97 [-1.65, -0.51]	3129.70 [3221.73, 2987.59]
ItierLab	215 – 275 ms	-4.53	41	< .001	-0.69 [-1.15, -0.32]	5623.22 [5404.37, 4699.71]
KHas	200 – 275 ms	-4.60	15	< .001	-1.09 [-2.17, -0.48]	192.78 [203.90, 195.63]
Krakow	195 – 260 ms	-11.59	25	< .001	-2.20 [-3.39, -1.53]	1.17e+09 [1.50e+09, 1.80e+09]
LSU	190 – 275 ms	-4.08	24	< .001	-0.79 [-1.47, -0.32]	364.85 [360.69, 323.03]
Magdeburg	190 – 260 ms	-7.56	24	< .001	-1.46 [-2.38, -0.91]	3.55e+05 [4.16e+05, 4.41e+05]
Malaga	200 – 265 ms	-6.39	27	< .001	-1.17 [-1.92, -0.69]	4.61e+04 [5.06e+04, 4.99e+04]
Munich	195 – 270 ms	-4.17	27	< .001	-0.77 [-1.39, -0.32]	209.00 [198.54, 171.69]
NCC_UGR	195 – 275 ms	-4.86	26	< .001	-0.91 [-1.59, -0.44]	1006.39 [1011.29, 917.90]
Neuruppin	195 – 265 ms	-3.81	26	< .001	-0.71 [-1.33, -0.26]	85.98 [79.71, 67.62]
ONERA	195 – 270 ms	-4.63	22	< .001	-0.93 [-1.70, -0.43]	436.67 [441.62, 404.00]
TrierCogPsy	190 – 270 ms	-4.58	11	< .001	-1.23 [-2.69, -0.52]	97.16 [105.88, 105.27]
TrierKamp	210 – 270 ms	-3.88	27	< .001	-0.71 [-1.32, -0.27]	104.27 [96.73, 82.07]
UNIMORE	190 – 265 ms	-8.80	19	< .001	-1.89 [-3.17, -1.20]	6.95e+05 [8.64e+05, 9.90e+05]
UniversityofVienna	185 – 255 ms	-3.98	23	< .001	-0.79 [-1.48, -0.30]	110.87 [105.89, 92.16]
Verona	185 – 275 ms	-6.15	26	< .001	-1.15 [-1.91, -0.66]	2.20e+04 [2.40e+04, 2.35e+04]
ZJU	190 – 290 ms	-3.24	26	< .001	-0.60 [-1.20, -0.16]	343.09 [333.49, 294.08]

Note. The p_{boot} values reported in this table reflect the median *p* values of the 1000 bootstrap procedures. Due to the way that the bootstrap procedure was implemented (see Methods section), some positive parametric *t* values resulted in significant p_{boot} values. Note that since we selected the time windows to include a negative component, in contrast to p_{boot} values, effect sizes and *BF*s for Colors and Forms were not bootstrapped and are therefore biased toward

negative values and evidence for the presence of a negative component, respectively; this bias does not apply to the Difference tests.

Meta-analysis

The random-effects meta-analytic estimate for Colors was $t(21) = -3.08$, $p = .005$ (see Figure 11), therefore this effect was replicated. For Forms, the estimate was $t(21) = -15.85$, $p < .001$ (see Figure 12), therefore this effect was replicated. For the difference between conditions, the estimate was $t(21) = -12.80$, $p < .001$ (see Figure 13), therefore this effect was replicated as well.

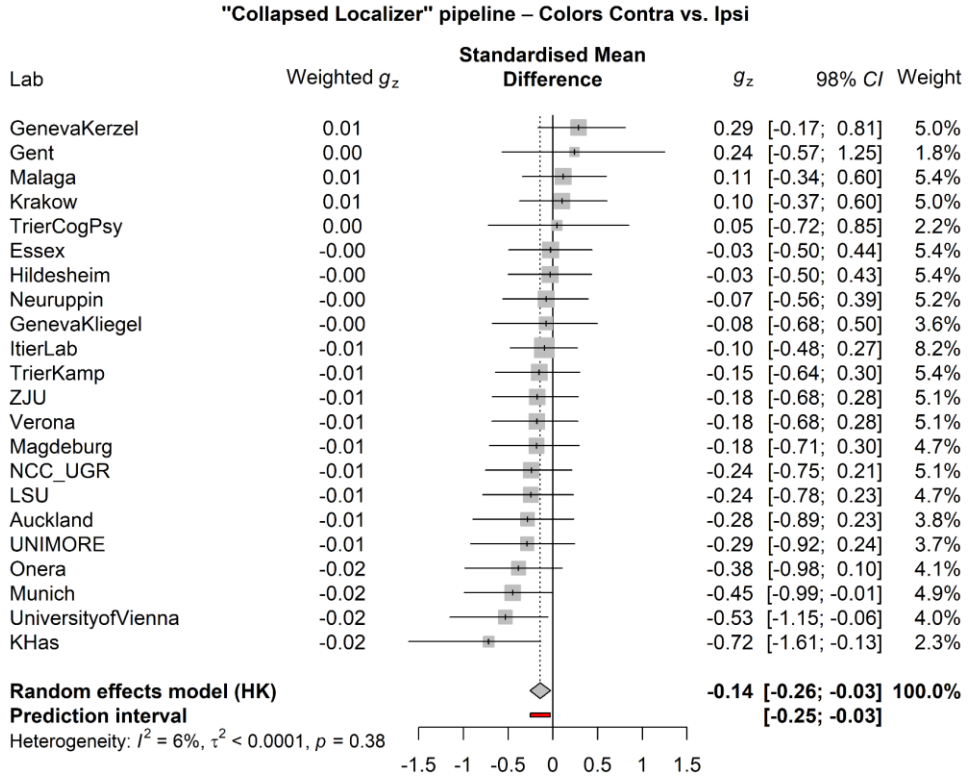


Figure 11. *Forest plot of the meta-analysis for Colors in the "Collapsed localizer" pipeline.*

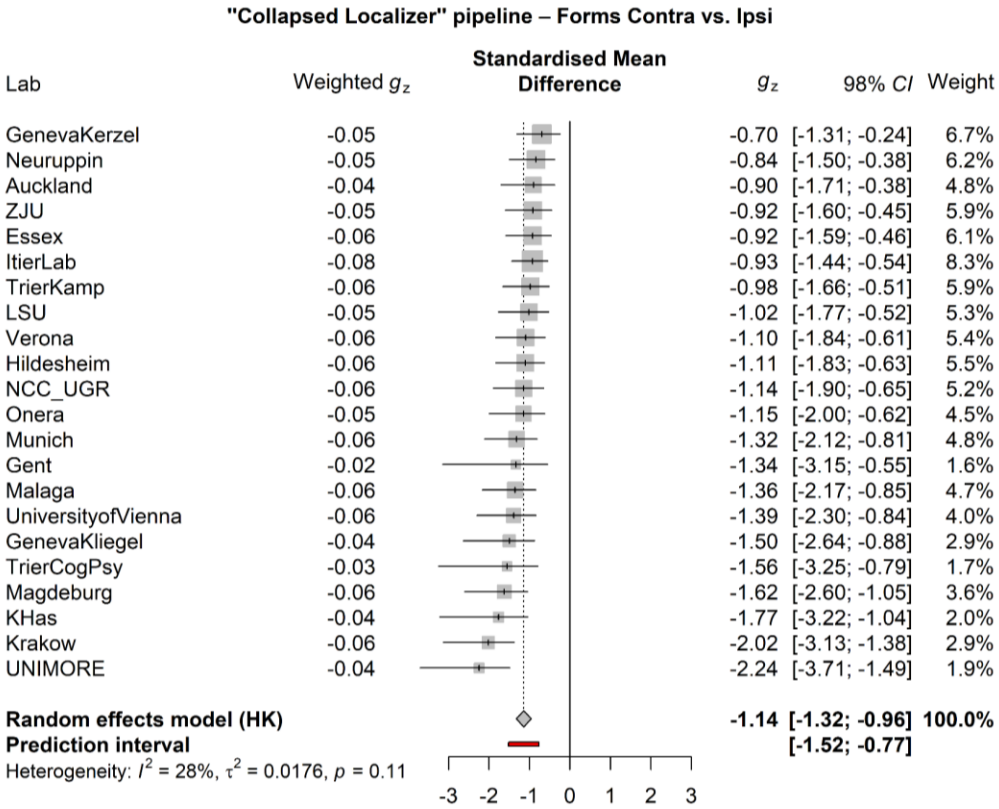


Figure 12. *Forest plot of the meta-analysis for Forms in the "Collapsed localizer" pipeline.*

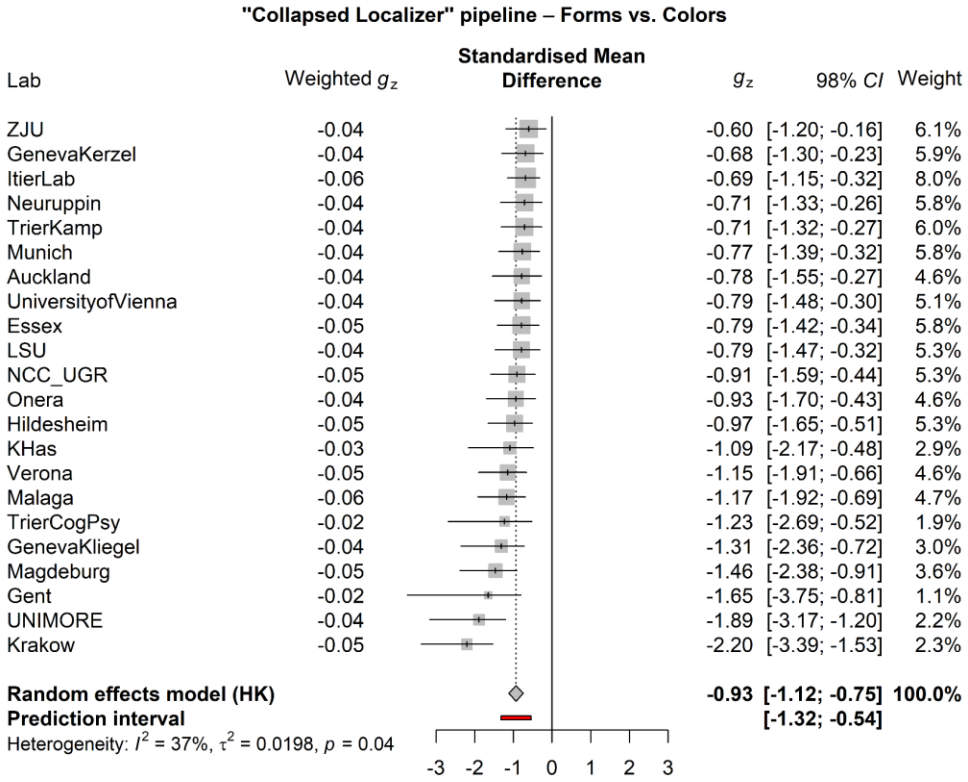


Figure 13. *Forest plot of the meta-analysis for Difference in the "Collapsed localizer" pipeline.*

ICA & Collapsed localizer pipeline

The color N2pc replicated in 16 labs out of 22 (see Table 5). The median g_z was -0.19. The form N2pc replicated in all labs. The median g_z was -1.18. The Difference between form and color N2pc replicated in all labs. The median g_z was -0.97.

Table 5. *Results from the ICA & Collapsed localizer pipeline.*

Lab	Time window	*t*	*df*	p_{boot}	g_z [98% *CI*]	BF_{-0} [wide, ultrawide]
			Colors			
Auckland	185 – 265 ms	-1.96	19	.001	-0.42 [-1.09, 0.10]	2.18 [1.78, 1.38]
Essex	200 – 275 ms	-0.21	27	.060	-0.04 [-0.51, 0.42]	0.24 [0.17, 0.13]

GenevaKerzel	195 – 255 ms	1.07	23	.075	0.21 [-0.28, 0.76]	0.11 [0.08, 0.06]
GenevaKliegel	195 – 265 ms	-0.91	15	.007	-0.21 [-0.92, 0.40]	0.58 [0.45, 0.33]
Gent	185 – 280 ms	0.23	8	.050	0.07 [-0.87, 1.07]	0.28 [0.21, 0.15]
Hildesheim	210 – 270 ms	-0.15	27	.077	-0.03 [-0.50, 0.44]	0.23 [0.17, 0.12]
ItierLab	215 – 275 ms	-0.80	40	.008	-0.12 [-0.51, 0.25]	0.36 [0.26, 0.19]
KHas	200 – 275 ms	-2.94	14	.002	-0.72 [-1.66, -0.12]	10.47 [9.66, 8.23]
Krakow	195 – 260 ms	0.06	25	.024	0.01 [-0.47, 0.50]	0.20 [0.14, 0.10]
LSU	185 – 270 ms	-1.76	21	.018	-0.36 [-0.97, 0.14]	1.58 [1.26, 0.96]
Magdeburg	190 – 255 ms	-1.74	23	.001	-0.34 [-0.92, 0.14]	1.49 [1.18, 0.89]
Malaga	200 – 265 ms	1.19	25	.017	0.23 [-0.24, 0.75]	0.10 [0.07, 0.05]
Munich	190 – 265 ms	-3.12	25	< .001	-0.59 [-1.20, -0.14]	18.59 [16.33, 13.30]
NCC_UGR	195 – 275 ms	-1.13	26	< .001	-0.21 [-0.72, 0.25]	0.62 [0.47, 0.34]
Neuruppin	195 – 265 ms	-0.41	26	.003	-0.08 [-0.56, 0.39]	0.29 [0.21, 0.15]
ONERA	195 – 270 ms	-1.86	22	< .001	-0.37 [-0.97, 0.11]	1.80 [1.44, 1.10]
TrierCogPsy	190 – 270 ms	0.12	11	.051	0.03 [-0.74, 0.83]	0.26 [0.20, 0.14]
TrierKamp	210 – 270 ms	-0.93	27	.009	-0.17 [-0.66, 0.28]	0.48 [0.36, 0.26]
UNIMORE	195 – 265 ms	-1.04	19	< .001	-0.22 [-0.84, 0.31]	0.62 [0.47, 0.35]
UniversityofVienna	185 – 255 ms	-2.52	23	< .001	-0.50 [-1.11, -0.02]	5.56 [4.68, 3.70]
Verona	180 – 270 ms	-1.01	25	.001	-0.19 [-0.71, 0.27]	0.55 [0.41, 0.30]

730

ZJU	190 – 290 ms	-1.05	26	< .001	-0.20 [-0.70, 0.26]	0.56 [0.42, 0.31]
			Forms			
Auckland	185 – 265 ms	-4.50	19	< .001	-0.97 [-1.83, -0.43]	252.00 [257.15, 237.65]
Essex	200 – 275 ms	-4.93	27	< .001	-0.91 [-1.57, -0.45]	1304.27 [1310.71, 1189.21]
GenevaKerzel	195 – 255 ms	-4.92	23	< .001	-0.97 [-1.73, -0.47]	898.53 [922.06, 854.27]
GenevaKliegel	195 – 265 ms	-6.52	15	< .001	-1.55 [-2.87, -0.87]	4534.23 [5340.49, 5742.00]
Gent	185 – 280 ms	-4.09	8	< .001	-1.23 [-3.14, -0.43]	29.67 [32.07, 31.85]
Hildesheim	210 – 270 ms	-6.02	27	< .001	-1.11 [-1.83, -0.63]	1.86e+04 [2.01e+04, 1.94e+04]
ItierLab	215 – 275 ms	-6.20	40	< .001	-0.95 [-1.48, -0.56]	1.27e+05 [1.31e+05, 1.20e+05]
KHas	200 – 275 ms	-7.67	14	< .001	-1.87 [-3.47, -1.11]	1.71e+04 [2.11e+04, 2.40e+04]
Krakow	195 – 260 ms	-11.99	25	< .001	-2.28 [-3.50, -1.59]	2.34e+09 [3.03e+09, 3.66e+09]
LSU	185 – 270 ms	-5.24	21	< .001	-1.08 [-1.92, -0.55]	1430.11 [1516.73, 1451.46]
Magdeburg	190 – 255 ms	-7.69	23	< .001	-1.52 [-2.48, -0.95]	3.51e+05 [4.15e+05, 4.46e+05]
Malaga	200 – 265 ms	-7.04	25	< .001	-1.34 [-2.18, -0.82]	1.44e+05 [1.64e+05, 1.69e+05]
Munich	190 – 265 ms	-7.93	25	< .001	-1.51 [-2.42, -0.96]	9.98e+05 [1.18e+06, 1.26e+06]
NCC_UGR	195 – 275 ms	-6.34	26	< .001	-1.18 [-1.95, -0.69]	3.44e+04 [3.79e+04, 3.75e+04]
Neuruppin	195 – 265 ms	-4.66	26	< .001	-0.87 [-1.54, -0.41]	634.25 [628.55, 563.56]
ONERA	195 – 270 ms	-5.83	22	< .001	-1.17 [-2.03, -0.64]	5799.53 [6336.15, 6241.74]
TrierCogPsy	190 – 270 ms	-5.64	11	< .001	-1.51 [-3.18, -0.76]	401.39 [465.96, 495.43]

TrierKamp	210 – 270 ms	-5.31	27	< .001	-0.98 [-1.66, -0.51]	3289.14 [3390.30, 3147.75]
UNIMORE	195 – 265 ms	-9.10	19	< .001	-1.95 [-3.27, -1.26]	1.14e+06 [1.42e+06, 1.64e+06]
UniversityofVienna	185 – 255 ms	-7.05	23	< .001	-1.39 [-2.31, -0.84]	9.43e+04 [1.09e+05, 1.13e+05]
Verona	180 – 270 ms	-6.15	25	< .001	-1.17 [-1.95, -0.67]	1.90e+04 [2.07e+04, 2.04e+04]
ZJU	190 – 290 ms	-4.74	26	< .001	-0.89 [-1.56, -0.42]	757.84 [755.11, 680.26]

Difference

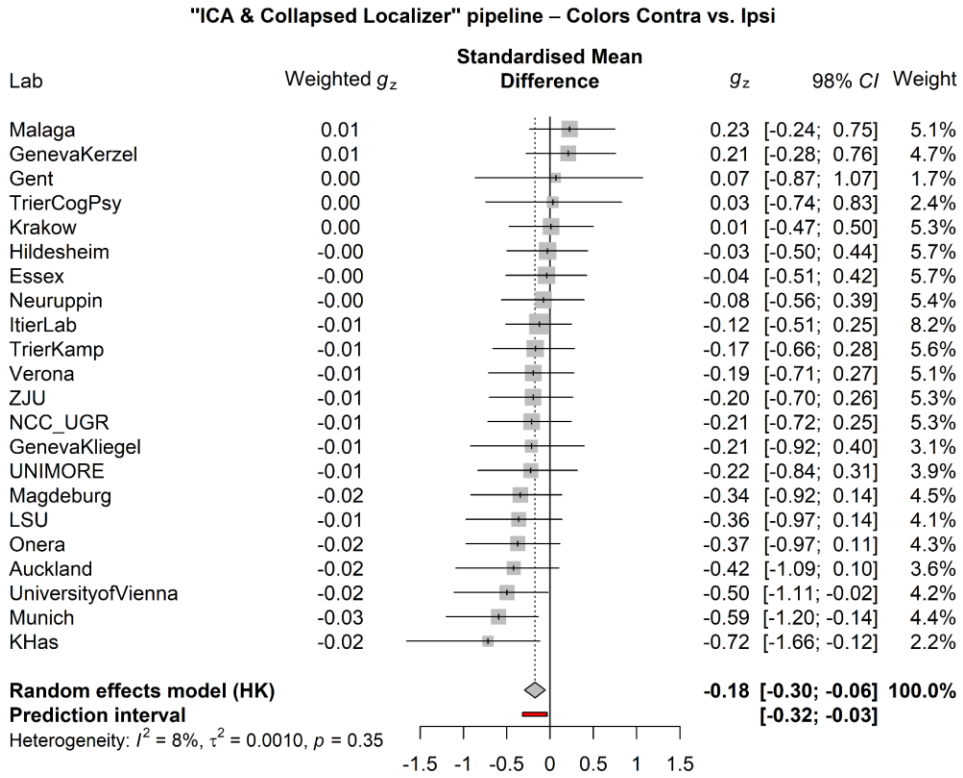
Auckland	185 – 265 ms	-3.96	19	< .001	-0.85 [-1.66, -0.32]	84.64 [82.69, 73.53]
Essex	200 – 275 ms	-4.20	27	< .001	-0.77 [-1.40, -0.32]	220.89 [210.22, 182.06]
GenevaKerzel	195 – 255 ms	-4.42	23	< .001	-0.87 [-1.60, -0.38]	294.72 [291.59, 261.46]
GenevaKliegel	195 – 265 ms	-4.96	15	< .001	-1.18 [-2.30, -0.56]	355.41 [385.26, 378.71]
Gent	185 – 280 ms	-4.39	8	< .001	-1.32 [-3.32, -0.51]	41.07 [45.34, 46.01]
Hildesheim	210 – 270 ms	-5.33	27	< .001	-0.98 [-1.66, -0.51]	3424.65 [3533.73, 3284.16]
ItierLab	215 – 275 ms	-4.65	40	< .001	-0.71 [-1.19, -0.34]	1230.17 [1146.88, 973.29]
KHas	200 – 275 ms	-4.84	14	< .001	-1.18 [-2.37, -0.54]	248.24 [269.02, 264.70]
Krakow	195 – 260 ms	-11.63	25	< .001	-2.21 [-3.40, -1.53]	1.25e+09 [1.60e+09, 1.92e+09]
LSU	185 – 270 ms	-4.05	21	< .001	-0.83 [-1.59, -0.33]	117.15 [113.93, 100.78]
Magdeburg	190 – 255 ms	-6.65	23	< .001	-1.31 [-2.20, -0.78]	4.08e+04 [4.63e+04, 4.74e+04]
Malaga	200 – 265 ms	-5.96	25	< .001	-1.13 [-1.91, -0.64]	1.22e+04 [1.32e+04, 1.29e+04]

Munich	190 – 265 ms	-5.40	25	< .001	-1.03 [-1.76, -0.54]	3329.40 [3488.22, 3291.71]
NCC_UGR	195 – 275 ms	-5.17	26	< .001	-0.97 [-1.66, -0.49]	2134.67 [2191.69, 2028.39]
Neuruppin	195 – 265 ms	-3.93	26	< .001	-0.73 [-1.36, -0.28]	112.68 [105.48, 90.17]
ONERA	195 – 270 ms	-4.72	22	< .001	-0.95 [-1.72, -0.44]	523.72 [532.90, 490.19]
TrierCogPsy	190 – 270 ms	-4.28	11	< .001	-1.15 [-2.55, -0.45]	63.75 [67.97, 66.12]
TrierKamp	210 – 270 ms	-3.81	27	< .001	-0.70 [-1.31, -0.26]	89.16 [82.27, 69.50]
UNIMORE	195 – 265 ms	-7.97	19	< .001	-1.71 [-2.90, -1.06]	1.74e+05 [2.12e+05, 2.35e+05]
UniversityofVienna	185 – 255 ms	-4.01	23	< .001	-0.79 [-1.49, -0.31]	119.23 [114.19, 99.60]
Verona	180 – 270 ms	-5.55	25	< .001	-1.06 [-1.80, -0.57]	4736.15 [5008.63, 4769.04]
ZJU	190 – 290 ms	-3.00	26	< .001	-0.56 [-1.14, -0.11]	14.67 [12.70, 10.23]

Note. The p_{boot} values reported in this table reflect the median *p* values of the 1000 bootstrap procedures.

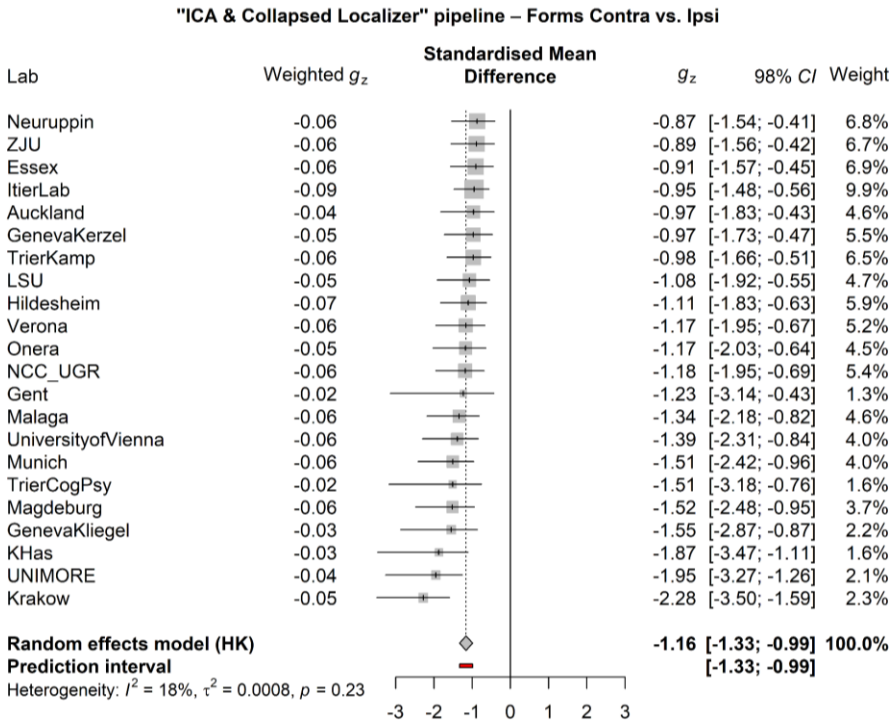
Meta-analysis

The random-effects meta-analytic estimate for Colors was $t(21) = -3.68$, $p = .001$ (see Figure 14), therefore this effect was replicated. For Forms, the estimate was $t(21) = -17.26$, $p < .001$ (see Figure 15), therefore this effect was replicated. For the difference between conditions, the estimate was $t(21) = -14.63$, $p < .001$ (see Figure 16), therefore this effect was replicated.



741

742 Figure 14. *Forest plot of the meta-analysis for Colors in the "ICA & Collapsed localizer" pipeline.*



743

744 Figure 15. *Forest plot of the meta-analysis for Forms in the "ICA & Collapsed localizer" pipeline.*

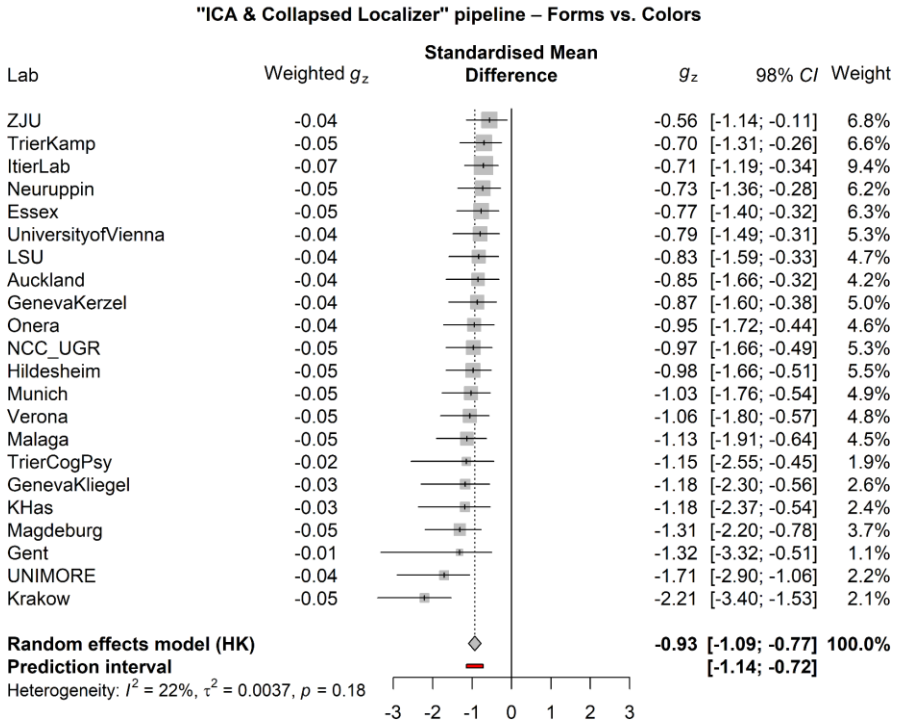


Figure 16. *Forest plot of the meta-analysis for Difference in the "ICA & Collapsed localizer" pipeline.*

Exploratory analyses with various time windows

The reported analyses are all based on the strong premise that the N2pc occurs in a fixed time window either across labs (original pipeline) or across conditions (collapsed localizer). This is a traditional assumption in the larger ERP literature, but may not necessarily be true. In fact, some would argue that it is highly unlikely that the cognitive processes (of which ERP components are purportedly an observable correlate) have a fixed timing independent of the stimuli and task (e.g., Liesefeld, 2018; Ouyang et al., 2011; Töllner et al., 2011). For the specific component of interest here, a rough review of the literature indicates that the amplitudes of components referred to as "N2pc" are measured in time windows that start as early as 140 ms (Papaioannou & Luck, 2020) up to as late as 350 ms (Woodman & Luck, 1999).

In practice, it is likely that most researchers investigating the N2pc do not determine their time windows a priori, but select the negativity from the difference wave that falls roughly into the commonly observed N2pc window. From our rough review of the N2pc literature, we thus found 17 different time windows. Some of these time windows are clearly stated as being created after visual inspection of the data, and for some it is plausible that they were based on visual inspection (especially when these windows are not consistently selected within a given lab). However, it is also worth noting that some labs have been very consistent across the years regarding the time window from which they extract the N2pc.

Table 6. *Number of labs replicating the N2pc (out of 22 labs in total) with various time windows found in the literature.*

Time window	Reference DOI	Condition	*N* (%) replicated	Average g_z
140 – 252 ms	10/gj6jd6	Colors	10 (45%)	-0.44
		Forms	22 (100%)	-1.04
		Difference	16 (72%)	-0.65
170 – 250 ms	10/fht828	Colors	16 (72%)	-0.57
		Forms	22 (100%)	-0.93

		Difference	12 (55%)	-0.50
175 – 325 ms	10/fskhpx	Colors	0 (0%)	0.29
		Forms	22 (100%)	-1.31
		Difference	22 (100%)	-1.35
180 – 235 ms	10/c69z2c	Colors	20 (91%)	-0.70
		Forms	14 (64%)	-0.63
		Difference	2 (9%)	-0.12
180 – 260 ms	10/b3s8s3	Colors	9 (41%)	-0.41
		Forms	22 (100%)	-1.11
		Difference	18 (82%)	-0.78
180 – 280 ms	10/d9whjn	Colors	3 (14%)	-0.09
		Forms	22 (100%)	-1.39
		Difference	22 (100%)	-1.19
	10/ghp3ng	Colors	1 (5%)	0.19

191 – 293 ms		Forms	22 (100%)	-1.47
		Difference	22 (100%)	-1.40
200 – 250 ms	10/cxvr7x	Colors	8 (36%)	-0.37
		Forms	22 (100%)	-0.96
		Difference	15 (38%)	-0.67
200 – 260 ms	10/fskhpx	Colors	4 (18%)	-0.22
		Forms	22 (100%)	-1.15
		Difference	21 (95%)	-0.93
200 – 275 ms	10/bj8mf5 10/ghp3ng 10/bc68bs	Colors	1 (5%)	0.03
		Forms	22 (100%)	-1.36
		Difference	22 (100%)	-1.23
200 – 280 ms	10/gj6bst 10/f4s98n	Colors	1 (5%)	0.11
		Forms	22 (100%)	-1.42
		Difference	22 (100%)	-1.31

200 – 300 ms	10/nhhc 10/gj6bh3 10/gc9mrs	Colors	0 (0%)	0.37
		Forms	22 (100%)	-1.48
		Difference	22 (100%)	-1.49
220 – 260 ms	10/fskhpx	Colors	0 (0%)	0.06
		Forms	22 (100%)	-1.24
		Difference	22 (100%)	-1.14
220 – 300 ms	**Original window**	**Colors**	**0 (0%)**	**0.61**
		Forms	**22 (100%)**	**-1.51**
		Difference	**22 (100%)**	**-1.61**
225 – 300 ms	10/grz7ps 10/d323p8	Colors	0 (0%)	0.67
		Forms	22 (100%)	-1.51
		Difference	22 (100%)	-1.64
235 – 290 ms	10/c69z2c	Colors	0 (0%)	0.68
		Forms	22 (100%)	-1.54

		Difference	22 (100%)	-1.64
260 – 360 ms	10/gc9mrs	Colors	0 (0%)	0.79
		Forms	22 (100%)	-0.90
		Difference	22 (100%)	-1.26
350 – 425 ms	10/bc68bs	Colors	1 (5%)	-0.16
		Forms	22 (100%)	-1.18
		Difference	22 (100%)	-1.16

We can see from Table 6 that with most time windows, the color N2pc still did not replicate. However, early time windows (ending at or before 250 ms) resulted in a significant N2pc to Colors for 36% to 91% of the labs. Interestingly, other studies with isolated stimuli (comparable to the present study) seem to be the ones that observed N2pcs in such an early time window (e.g., Brisson et al., 2007; Papaioannou & Luck, 2020).

Exploratory results – Behavioral measures

As this will be of interest to some readers, we additionally report analyses on reaction times and error rates. For the reaction time analyses, we extracted reaction times from correct

trials with distractors (i.e., excluding the target-only trials) that were not rejected for eye-movement artifacts in the "Original" pipeline. We computed a two-sided paired-samples *t* test between the average reaction times of the two conditions for each lab. There was a significant difference in all labs. We then computed a meta-analytic *p* value and effect size with the same procedure as the one used for the ERP analyses, $t(21) = 18.31$, $p < .001$, $g_z = 1.34$ [1.15, 1.52]. On average (pulling together the data from all participants), participants were faster for Colors than for Forms (481 ms vs. 555 ms; within-subject 98% *CI*: 3.83 ms; see Figure 17).

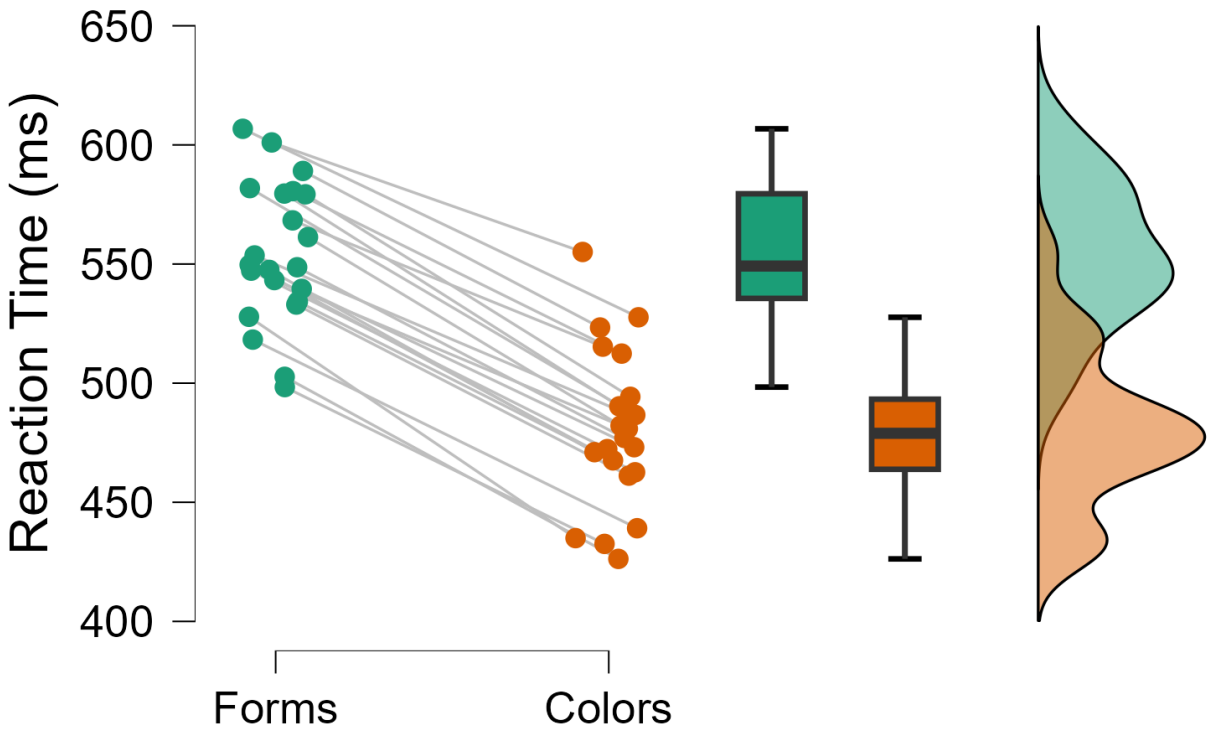


Figure 17. *Results from the exploratory reaction time analysis.*

Note. Each dot represents the average reaction time of all participants from a given lab in the respective distractor ~~a given~~ condition. Reaction times from correct trials that were not rejected in the "Original" pipeline were used.

We also analyzed the accuracy in each condition. For this analysis, we used the same procedure, except that we kept incorrect trials and trials rejected due to eye-behavior. There was a significant difference in only 9 out of 22 labs. However, given the meta-analytic *p* value and effect size we still conclude that there was an effect on error rates, $t(21) = 9.46$, $p < .001$, $g_z = 0.41$ [0.30, 0.52]. On average (pulling together the data from all participants), participants were better for Colors than for Forms (94.41% vs. 92.79%; within-subject 98% *CI*: 0.30%; see Figure 18).

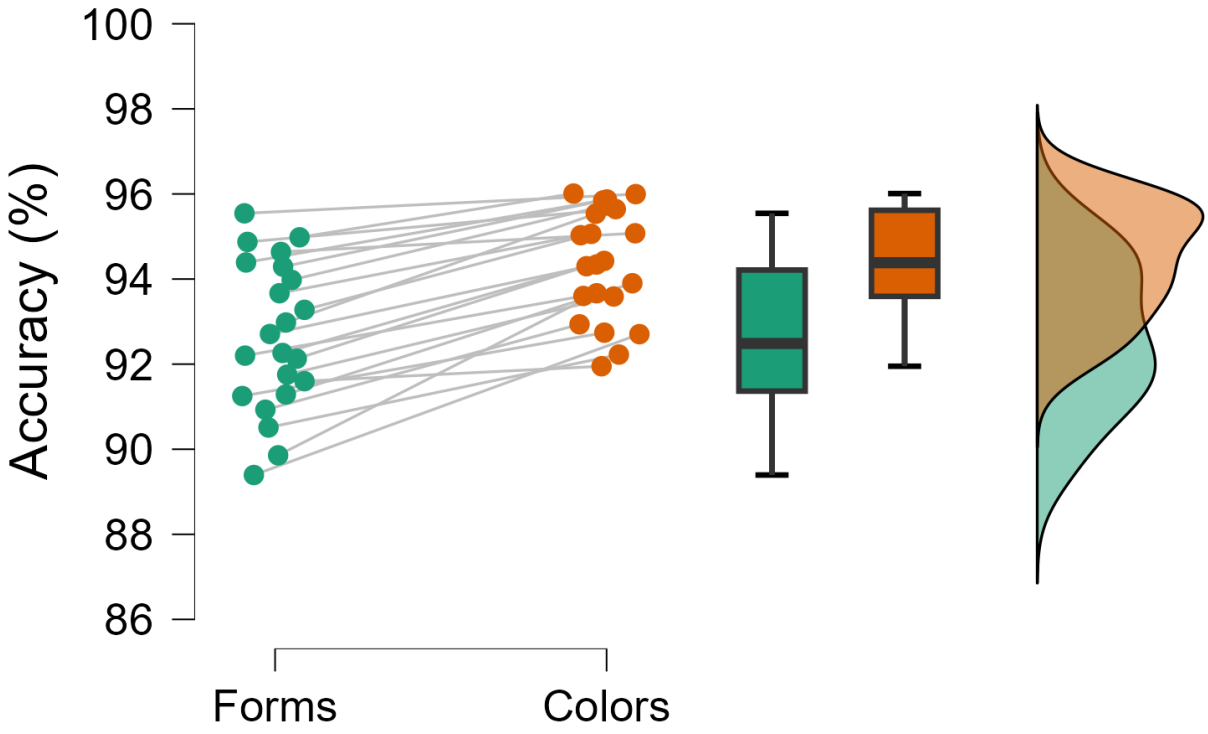


Figure 18. *Results from the exploratory response accuracy analysis*.

Note. Each dot represents the average accuracy of all participants from a given lab in the respective distractor ~~a given~~ condition.

Exploratory analyses – Less strict trial rejection criteria

Most labs ended up sampling more than the initial 28 participants because the trial rejection (and subsequent participant rejection) criteria were quite strict. The rather high exclusion rate is likely due to the fact that the replicated search window for artifacts was overly wide and we therefore lost too many trials. In particular, trials were flagged as contaminated if there were any eye-movements or blinks at any point during the trial (i.e., from -100 to +600 ms relative to display onset). This time window is likely too wide given that we focused our analyses on the 220 – 300 ms time window. Rejecting trials due to eye-related behavior happening during or even after the N2pc time window seems too strict, because the perceptual input eliciting the N2pc already disappeared (after 150 ms). Indeed, of these 241 excluded participants, 123 (51%) had most trials rejected due to blinks, 109 (45%) because of eye movements, and only 9 (4%) because they made too many mistakes in the task. If we pull together the 241 rejected participants from the original pipeline, the pattern of results is overall very comparable to that of non-rejected participants (see Figure 19).

In the present exploratory analysis, hereafter called the "Less Strict" pipeline, we slightly modified the “Original” pipeline to restrict the search window for blinks and eye-movements to -100 – +150 ms. With this narrower window, 10 participants were excluded because their HEOG in the lateralized ERP exceeded our threshold, while only one participant was excluded for this reason with the original search window. The first consequence was a large increase in the number of trials per condition for each participant. The average number of rejected trials (for non-rejected participants) for Colors and Forms went from 29.54% and 33.29% in the “Original” pipeline to 11.63% and 13.43% in the “Less strict” pipeline. In other words, this added on average 47 and 52 trials to each ERP.

To quantify the effect of this, in both pipelines, for each participant in both conditions, we computed 100 bootstrapped standard measurement errors (bSME; 1000 iterations; Luck et al., 2021) and kept the median value of these 100 bootstrap procedures. We used the 170 - 250 ms time-window because it captures both the color N2pc and most of the form N2pc. As nine additional participants were rejected from the less-strict pipeline due to the HEOG criterion, we included data from the 529 participants common to both pipelines. In both conditions, the bSME of 486 participants (91.8%) was improved in the Less-strict compared to the Original pipeline. There were 18 participants for whom the bSME improved for Forms but worsened for Colors, and another 18 with the opposite pattern. This leaves only 7 participants (1.3%) who ended up with a decrease in data quality in the less-strict pipeline. The average bSME improvement over these 529 participants was 14.6% for Colors and 12.5% for Forms.

For each lab, we then computed the root mean square (RMS) of the bSME of each participant (on all participants accepted in the Less-strict pipeline on the one hand and all participants from the Original pipeline on the other hand). The median RMS(bSME) for Colors were at 0.408 and 0.462 in the Less strict and Original pipelines respectively. For Forms they were at 0.399 and 0.458. The median of the differences were 9.8% and 6.3% higher (worse) in the Original pipeline. To note, we report the median because, while the RMS(bSME) improved for most labs, there were some labs for which it actually got considerably worse in one or both conditions.

The indirect consequence of the narrow artifact-search window was that far fewer participants were rejected due to an insufficient number of trials. Indeed, with the narrow window, only 13 participants were rejected due to that criterion compared to 241 before. The overall number of excluded participants was 37, which means that the number of valid

847 participants totaled at 742 participants. To test how this change in sample size affected our
848 results while also taking the effect of including potentially noisier data, we applied the following
849 procedure:

850 1. On the difference waves from the "Original" pipeline, we computed a meta-
851 analysis with the means extracted from the 170 - 250 ms time window (in which
852 16 labs had replicated the color N2pc). This allowed us to get more meaningful
853 comparisons of post-hoc power for Forms (in the original 220 - 300 ms time
854 window, power was virtually at 100% for all labs). This analysis window also
855 captures part of what we tentatively interpret as the color N2pc.
856 2. For each condition, we then computed the post-hoc power (one-sided, $\alpha = .02$) of
857 each lab using the meta-analytical effect size. The effect-size estimate was
858 therefore fixed between labs. We used this one rather than the mean or median
859 effect size across labs because it better represents the "true" effect size (i.e., this is
860 the one people would use in a power analysis to determine sample size) and is less
861 prone to random variations caused by low sample size.
862 3. We repeated steps 1. and 2. in the "Less strict" pipeline, using its meta-analytical
863 effect sizes.

864 This resulted in an average increase in power of 12.23% for Colors, 5.71% for Forms and
865 19.75% for the Difference between Forms and Colors. Notably, the power for Colors *increased*
866 despite the effect size being *smaller* in the less strict pipeline (see Table 7).

867 **Table 7.** *Effect sizes and power in the Original and Less strict pipelines.*

Condition	Meta Effect size Original	Meta Effect size Less Strict	Average Power Original	Average Power Less Strict
Colors	0.514	0.493	62.89%	75.12%
Forms	0.836	0.886	93.89%	99.61%
Difference	0.466	0.490	54.81%	74.56%

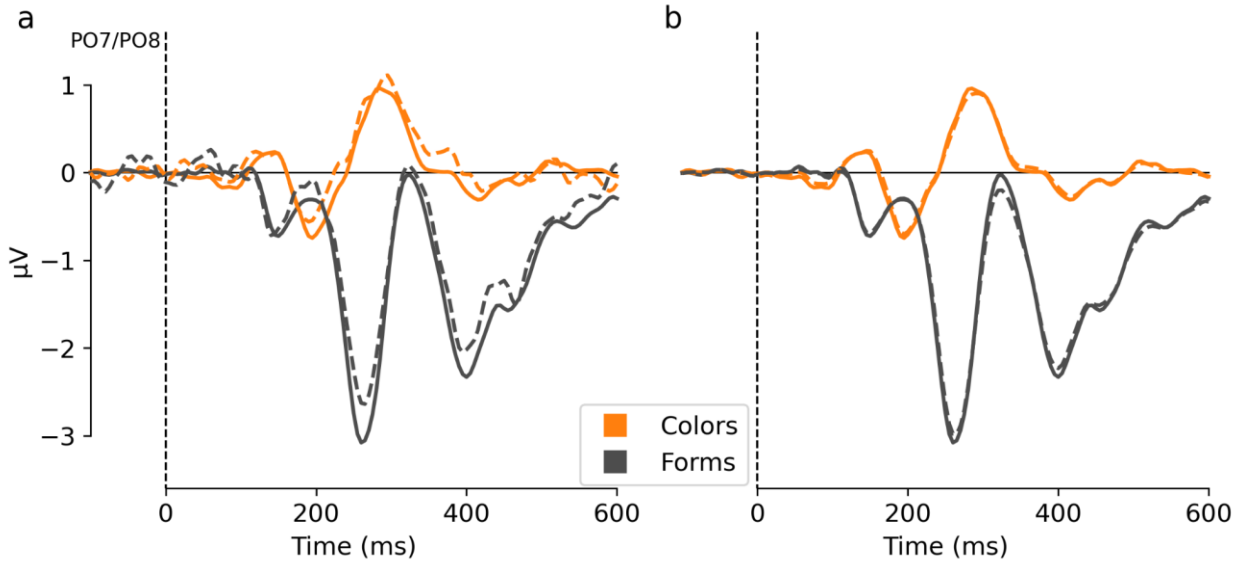


Figure 19. *Comparison of the ERPs depending on the rejection criteria*

Note. a) Comparison of rejected (full line) vs. non-rejected (dashed line) participants in the Original pipeline. b) Comparison of the Original pipeline (full line) with the Less-strict pipeline (i.e., rejected participants combined with non-rejected ones; dashed line).

Discussion

When we started this project, we felt very confident that we could replicate the highly influential N2pc results of Eimer (1996). After all, the N2pc has been observed in countless

studies and is a core tool in neurocognitive research on visual attention. This is also reflected in the outcome of the prediction markets conducted within the scope of our encompassing #EEGManyLabs project; on a scale from 0.00 to 1.00, researchers rated the likelihood of our replication attempt being successful at 0.906. We successfully replicated the form N2pc indeed. Yet, according to the pre-planned criteria and current standards, we did not replicate the color N2pc using the original pipeline. ~~This non-replicated result is arguably the more influential of the two, because far more N2pc studies use color patches than line patterns (W and M) as stimuli.~~ However, across the 22 replication attempts of the present study, ERP patterns were stunningly consistent for both conditions (see Figure 20), providing empirical evidence for the high quality and feasibility of the #EEGManyLabs approach.

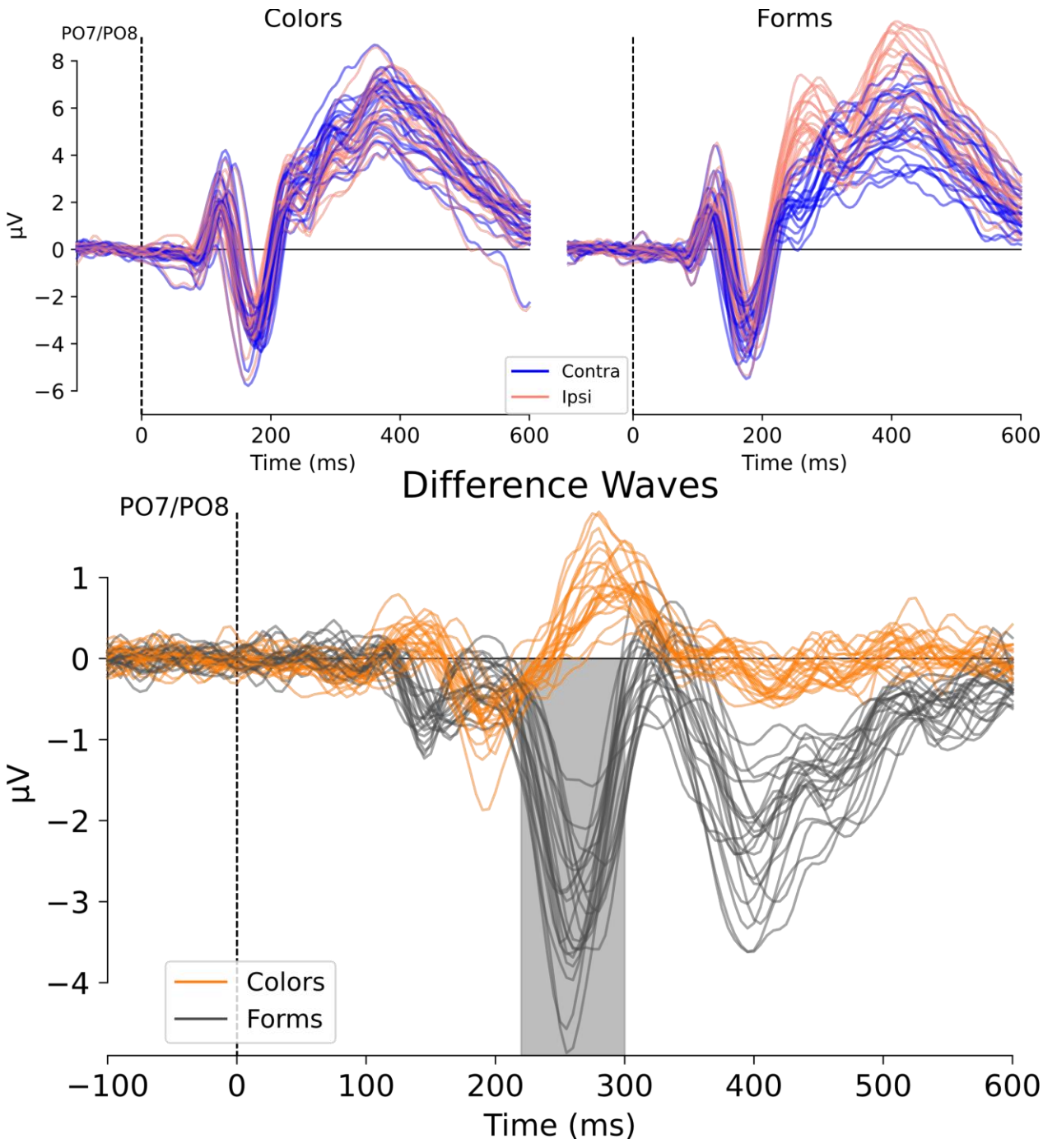


Figure 20. *Grand average waveforms from each lab*.

Note. Each individual line represents the grand average waveform from one lab in a given condition in the Original pipeline. Top panels: Contra- and Ipsi-lateral waveforms for both conditions. Bottom panel: Contra *minus* ipsi difference waveforms.

Visual inspection of the lateralized ERPs as well as our exploratory analyses might indicate that one reason for the highly consistent non-replication was that the component that could be classified as the color N2pc occurred in a different-than-expected time window.[1] The color N2pc was significant for 16 labs in our pre-registered collapsed-localizer pipeline and for 20 labs in one of our exploratory analyses using a different time window taken from the N2pc literature. This time window was not expected based on the original Eimer (1996) study, but could have been (approximately) expected based on other studies using sparse search displays (e.g., Brisson et al., 2007; Papaioannou & Luck, 2020). Despite its name, the N2pc is not tied in any way to the N2 component of the ERP - it might merely have happened to occur in this time range in the task design in which it has been discovered and therefore originally showed up as a modulation of the N2 (increased N2 at the contra- compared to ipsilateral electrode sites). In fact, in our data, there is not even a pronounced N2 in the ERP. As a consequence, there is no strict rule to select an analysis window for this component. Our choice of analysis window was based on the original study in our "Original" and "ICA" pipelines and on a pooling approach in our collapsed-localizer pipelines. The reconstructed lateralized ERPs (which were not shown in the original study) had already indicated that the N2pc occurs at different time points in the two conditions (and we preregistered an adapted collapsed localizer approach accordingly).

One potential reason for why this - now so obvious - latency difference between color and form N2pc (with a difference in peak latencies of 25 ms in the original study and 65 ms in the replication attempt) might have not been discovered and highlighted in the original study is a

[1] The other reason is that the color N2pc is rather small in amplitude. As pointed out by Martin Eimer (personal communication, February 17, 2025) it is much smaller than the N2pc to comparable color stimuli later measured by his team (Grubert & Eimer, 2013, 2015). This might have to do with the fact that color acted as search-guiding and reported feature in the present study, whereas it acted merely as a search-guiding feature and participants reported another feature of the stimulus in the Grubert and Eimer (2013, 2015) studies (see Liesefeld et al., 2024, for the distinction).

conviction ingrained in the ERP community: ERP components supposedly have a fixed timing, so that a given component should be measured in the same analysis window across conditions and studies. This likely stems from the practice in the early days of ERP research to name components by their timing (in addition to their polarity and topography). While the fixed-timing assumption ~~this~~ has been challenged (e.g., Liesefeld, 2018; Ouyang et al., 2011~~; Töllner et al., 2011~~) and despite early reports of variation in component latency (Kutas et al., 1977; Polich, 1987), including the N2pc (e.g., Hickey et al., 2010; Töllner et al., 2011; Woodman & Luck, 1999), the belief ~~in a fixed component timing~~ that a specific component occurs in a relatively narrow, fixed time interval is still widely held. One prominent consequence of this belief is the advice to analyze ERP components in a fixed time window that is ideally predetermined or, alternatively, based on a collapsed localizer (see Kappenman & Luck, 2016; Luck & Gaspelin, 2017). Strictly following this advice (as done here) can result in analysis windows that miss the component of interest, capture only part of this component or span several components. All three cases are nicely exemplified in the present study (see Figures 1d, 3 and 7): (a) by using the original N2pc analysis window (across studies), we almost completely missed what can be interpreted as the color N2pc; (b) by using the same window for both conditions, Eimer (1996) as well as some of our collapsed-localizer windows captured only part of the form N2pc; (c) most of the windows resulting from the collapsed localizer approach span the color N2pc and the ensuing positivity in our replication attempts. Thus, instead of considering the color N2pc as non-replicated, an alternative interpretation of this failed replication attempt might be that the belief that a given component has a constant timing with respect to an external event, independent of the exact circumstances under which it emerges, misleads ERP research and should be put to rest. The differences in component timing between the original study and our

replication attempt together with the high consistency across labs indicates that we did not exactly replicate all relevant parameters affecting the components' latencies. As the relevant information is no longer available, we can only speculate on some possible deviations in the following.

The delay in the ~~form~~ N2pc of the original study (relative to our 22 replication attempts) could be explained by a delay between the recorded marker time and the stimuli's appearance on screen in the original study.[2] We actually encountered this situation with a lab participating in the present replication study. Their ~~form~~ N2pcs seemed delayed compared to the other labs and their form N2pc was actually replicating almost perfectly the one that Eimer (1996) had found. We thus asked them to measure with a photodiode the delay between marker onset and stimuli's onset. They measured an average delay of approximately 40 ms. After correcting this delay, their data were much more coherent with that from the other labs (and thus less similar to Eimer's data).

In an attempt to gauge the delay that might have been induced by (compared to current standards) outdated hardware in the original study, we compared the peak latencies of the exogenous P1 and N1 ERP components. These were 12.5 ms and 10 ms shorter, respectively, in our replication attempt than in the (reconstructed) original data (see caption of Figure 2 for details). This represents less than a single frame at the 60 Hz display refresh rate presumably used in the original study.[3] Such slightly shorter latencies of the exogenous components might be

[2] Checking stimulus timing with photodiodes, as well as luminance measurement (see below), became a standard procedure in the Eimer lab only later (Martin Eimer, personal communication, February 17, 2025).

[3] This refresh rate is our best guess based on the faint memory of one co-author (AW) who contributed as a student assistant to the original study. This guess is supported by a published paper on another study conducted around the same time in the same lab, which reports a 60-Hz refresh rate (Eimer & Schlaghecken, 1998).

expected for two reasons: (1) 9 of the 22 contributing labs used display refresh rates higher than 60 Hz (stimuli at the vertical center of the display will appear approximately 4 ms earlier on a 120 Hz display than on a 60 Hz display relative to a marker at screen flip). (2) All contributing labs used considerably higher sampling rates (≥ 500 Hz), which allowed for higher cutoff frequencies of the online low-pass (antialiasing) filter (the low cutoff frequency online low-pass filter in the original study potentially may have introduced small delays into the signal; in contrast to the zero-phase filter used here for offline low-pass filtering and downsampling). Therefore, we assume that the delays between marker and stimulus onset were small and comparable between the original study and our replication attempt.

In any case, these slight delays cannot explain the considerably shorter N2pc peak latencies in our replication attempt. Compared to the (reconstructed) original data our N2pcs peaked 30 ms earlier for Forms (260 vs. 290 ms) and 70 ms earlier for Colors (195 vs. 265 ms; assuming that the earlier negative deflection in the difference wave indeed is a color N2pc). In contrast, reaction times in the replication were slower than in the original study by 48 ms for Forms (555 vs. 507 ms) and by 13 ms for Colors (481 vs. 468 ms). The overall slower reaction times in the replication may indicate differences in the speed-accuracy trade-off (unfortunately, accuracy was not reported in the original manuscript) due to differences in instruction and feedback, population, or other unknown differences (Heitz, 2014; Wickelgren, 1977).

A plausible explanation for the particularly large difference in timing of the color N2pc in the original Eimer (1996) study and our 22 replication attempts would be a difference in the displayed colors: color settings employed here reflected only the best guess of the original author, because the original experimental program had been lost and colors were not measured. Even when the experimental program is available for a replication study, colors are typically

specified in the RGB colorspace or a linear transformation thereof such as HSV (only providing information about how much each sub-pixel is stimulated, but not what the resulting color is), which means one can only know the approximate chromaticity of the colors and there's no information about their absolute luminance. Furthermore, employed monitors are often not calibrated and objective color measurements are rarely performed. However, variation induced by non-calibration cannot have had a huge effect, because otherwise the pattern would not be so consistent across replicating labs (Figure 20). A systematic difference between original and replication studies might be that screens were generally dimmer at the time when the original Eimer study was conducted.[4]

Whatever the source of the potential variation in color, as N2pc timing depends on stimulus salience (Töllner et al., 2011) and salience of the color patches would depend on the color-to-background contrast (including the luminance difference), it appears likely that the colors in the original Eimer (1996) study were less salient. Notably, this speculation would not only explain why the original study observed a relatively late color N2pc, but it would also explain why the latency-difference between the two N2pcs was smaller in the original study compared to most of the replication results reported here: a decrease in contrast should have a weaker effect on salience of the high-contrast white letters on a gray background in Forms compared to salience of the color patches in Colors. The thereby induced similarity in latency of the two N2pcs had allowed Eimer to observe them in the same time window (which matches the weaker color N2pc better than the stronger form N2pc as evident in Figure 1b, though). If there

[4] We thank Clayton Hickey for pointing this out to us.

had not been a much larger difference in timing between the two N2pcs, replication rate in our collapsed localizer pipelines would have been much higher.

In general, the comparison of N2pc peak latencies between the two studies demonstrates the variability of the timing of ERP components and their sensitivity to small differences (which we had hoped to avoid in our replication attempt). A lesson that can be learned from this observation is that, for replication attempts of EEG patterns, the exact stimulation is of higher importance than for replication attempts of purely behavioral studies. Unfortunately, it is hardly if ever possible to exactly reproduce the original stimulation due to differences in hardware and incomplete reporting of stimulation parameters (e.g., the actually produced colors). This may prove to be a major obstacle for the replication of ERP studies, especially when the original studies were conducted long ago, and some crucial information on the exact recording and stimulation parameters is missing. This difficulty can be circumvented to a certain degree, by anticipating potential differences in component latency in future replication attempts. A recent paper from Lepauvre et al. (2024) advises measuring marker-to-display onset latency. Based on our experience with the present replication project, we agree that this is indeed an important step in EEG research. We would also add that measuring and reporting colors in xyY (or XYZ) coordinates is important, as this would allow replications to get much closer to the exact stimulation, which could impact replicability. This can be achieved with a reasonable precision using consumer-grade hardware that can be acquired for less than 200€ and operated with open-source software.

Future use of our massive data set

Given its substantial size (779 full datasets; 264 trials for each participant in each relevant condition; before any trial or participant exclusion), the present data set might be of use

to study further questions related to the N2pc, the extraction of (lateralized) ERPs, and other analysis techniques (e.g., time-frequency-analyses or decoding approaches). As an example, we compared N2pc results for rejected and non-rejected data sets and evaluated the analysis decision to exclude trials with artifacts in a wide search window. It turned out that results were highly comparable for rejected participants and that a narrower artifact-search window could increase the power to detect effects. It would be interesting to examine how other analysis decisions affected the power or other metrics of data quality.

Another issue to address is the question on the relation between the N2pc and behavioral (or attentional) performance, thereby on possible functional interpretations of the N2pc. For instance, does a higher individual N2pc amplitude indicate a more or less efficient deployment of attention? Assuming that a larger N2pc indicates a stronger involvement of the selection mechanism (e.g., Luck et al., 1997; Śmigasiewicz et al., 2015), we might expect that the N2pc amplitude is positively correlated with behavioral efficiency (the larger the N2pc, the faster the RTs and the lower the error rates). On the other hand, based on the same assumption, the current observation of larger amplitude and delayed latency of the N2pc in Forms compared to Colors (and the corresponding RT and accuracy condition differences) might be compatible with findings suggesting that the N2pc is related to selection difficulty, and not to selection efficiency. For example, Asanowicz et al. (2021) observed that in the flanker task, the N2pc was larger in the perceptually more difficult incongruent flanker condition than in the congruent condition. The N2pc amplitude was positively correlated with the behavioral cost of flanker interference, with larger N2pcs indicating a less efficient behavioral performance (specifically, the incongruent – congruent difference in N2pc amplitudes correlated positively with the incongruent – congruent difference in RTs). Thus, a larger N2pc could be related to perceptual

difficulty and thereby to the "need" for selection. In other words, rather than a more efficient attentional processing, a larger N2pc could reflect a more effortful one.

Conclusion

Across all labs and analysis pipelines, we successfully replicated Eimer (1996)'s form N2pc. While our replication attempt technically failed for Eimer's color N2pc, we do not think that this demonstrates that the color N2pc was due to serendipity. Rather, our replication study highlights weaknesses in previous EEG research that can be ameliorated by more careful measurement and reporting of timing and stimulation (color in particular) and by improvements in analysis approaches and the underlying basic assumptions. Furthermore, our comparison of ERPs for "valid" and rejected datasets indicates that overly conservative rejection criteria do more harm than good by scrapping perfectly valid data. Most importantly, future (replication) studies should take into account that there is genuine variability in ERP component latency as one should expect if these components are correlates of temporally variable cognitive processes. Thus, our "failure" to exactly replicate Eimer's color N2pc can serve as a useful warning for future EEG replication attempts: component latency hinges on many influences, some of which are likely overseen or no longer reconstructable during replication. As a consequence, the chosen analysis windows might miss the component of interest. Our hope is that the present massive data set will generate even more insights on the N2pc and ERP methods in general.

References

Asanowicz, D., Panek, B., & Kotlewska, I. (2021). Selection for action: The medial frontal cortex is an executive hub for stimulus and response selection. *Journal of Cognitive Neuroscience*, *33*(8), 1442–1469. https://doi.org/10/g8vqr8

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

Brisson, B., Robitaille, N., & Jolicœur, P. (2007). Stimulus intensity affects the latency but not the amplitude of the N2pc. *NeuroReport*, *18*(15), 1627. https://doi.org/10/fht828

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

Constant, M. (2025). *Martin2Constant/EEGManyLabs_Eimer1996: 1.0* (Version 1.0) [Computer software]. Zenodo. https://doi.org/10/n3rg

Constant, M., Mandal, A., Asanowicz, D., Panek, B., Kotlewska, I., Yamaguchi, M., Gillmeister, H., Kerzel, D., Luque, D., Molinero, S., Vázquez-Millán, A., Pesciarelli, F., Borelli, E., Ramzaoui, H., Beck, M., Somon, B., Desantis, A., Castellanos, M. C., Martín-Arévalo, E., … Liesefeld, H. R. (2025a). *EEGManyLabs—Replication Raw Dataset—Eimer1996* (Version 1.0) [Dataset]. G-Node. https://doi.org/10/pmg4

Constant, M., Mandal, A., Asanowicz, D., Panek, B., Kotlewska, I., Yamaguchi, M., Gillmeister, H., Kerzel, D., Luque, D., Molinero, S., Vázquez-Millán, A., Pesciarelli, F., Borelli, E., Ramzaoui, H., Beck, M., Somon, B., Desantis, A., Castellanos, M. C., Martín-Arévalo, E., … Liesefeld, H. R. (2025b). *EEGManyLabs—Replication Processed Dataset—Eimer1996* (Version 1.0) [Dataset]. G-Node. https://doi.org/10/pmg5

Constant, M., Mandal, A., Asanowicz, D., Yamaguchi, M., Gillmeister, H., Kerzel, D., Luque, D., Pesciarelli, F., Fehr, T., Mushtaq, F., Pavlov, Y. G., & Liesefeld, H. R. (2023). A multilab investigation into the N2pc as an indicator of attentional selectivity: Direct replication of Eimer (1996). *PsyArXiv*. https://doi.org/10/n6xg

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

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1737–1747. https://doi.org/10/bvcjmn

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

Grubert, A., & Eimer, M. (2013). Qualitative differences in the guidance of attention during single-color and multiple-color visual search: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1433–1442. https://doi.org/10/f5c33f

Grubert, A., & Eimer, M. (2015). Rapid parallel attentional target selection in single-color and multiple-color visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 86–101. https://doi.org/10/f6x4t6

Hedges, L. V. (1981). 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. https://www.sciencedirect.com/book/9780080570655/statistical-methods-for-meta-analysis

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*. https://doi.org/10/gfw8p2

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

Hickey, C., van Zoest, W., & Theeuwes, J. (2010). The time course of exogenous and endogenous control of covert attention. *Experimental Brain Research*, *201*(4), 789–796. https://doi.org/10/fv8xm4

Hilimire, M. R., Hickey, C., & Corballis, P. M. (2012). Target resolution in visual search involves the direct suppression of distractors: Evidence from electrophysiology. *Psychophysiology*, *49*(4), 504–509. https://doi.org/10/cmvdph

IntHout, 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

JASP Team. (2024). *JASP* (0.19.1) [Computer software]. https://jasp-stats.org/

Kappenman, E. S., & Luck, S. J. (2016). Best practices for Event-Related Potential research in clinical populations. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(2), 110–115. https://doi.org/10/gfz986

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Supplement). 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

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792–795. https://doi.org/10/ctnwm4

Lepauvre, A., Hirschhorn, R., Bendtz, K., Mudrik, L., & Melloni, L. (2024). A standardized framework to test event-based experiments. *Behavior Research Methods*. https://doi.org/10/g8npf

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. (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., Constant, M., & Pavlov, Y. G. (2022). *Replication of Eimer (1996)—Experiment 2*. https://doi.org/10/n6xh

Liesefeld, H. R., Lamy, D., Gaspelin, N., Geng, J. J., Kerzel, D., Schall, J. D., Allen, H. A., Anderson, B. A., Boettcher, S., Busch, N. A., Carlisle, N. B., Colonius, H., Draschkow, D., Egeth, H., Leber, A. B., Müller, H. J., Röer, J. P., Schubö, A., Slagter, H. A., … Wolfe, J. (2024). Terms of debate: Consensus definitions to guide the scientific discourse on visual distraction. *Attention, Perception, & Psychophysics*. https://doi.org/10/gtcj55

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., 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

Liesefeld, H. R., & Müller, H. J. (2021). Modulations of saliency signals at two hierarchical levels of priority computation revealed by spatial statistical distractor learning. *Journal of Experimental Psychology: General*, *150*(4), 710–728. https://doi.org/10/gqdh98

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

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). **JASP**: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2). https://doi.org/10/ggbnjf

Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In E. S. Kappenman & S. J. Luck

(Eds.), *The Oxford handbook of event-related potential components* (pp. 330–360). 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

Luck, S. J., Stewart, A. X., Simmons, A. M., & Rhemtulla, M. (2021). Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology*, *58*(6). https://doi.org/10/gj6bjk

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). *markummitchell/engauge-digitizer: Version 12.1 Directory dialogs start in saved paths* (v12.1) [Computer software]. Zenodo. 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

Morey, R. D., & Rouder, J. N. (2024). *BayesFactor: Computation of Bayes Factors for common designs* (Version 0.9.12-4.7) [Computer software]. https://CRAN.R-project.org/package=BayesFactor

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/bf6sxj

Papaioannou, O., & Luck, S. J. (2020). Effects of eccentricity on the attention-related N2pc component of the event-related potential waveform. *Psychophysiology*, *57*(5). https://doi.org/10/gj6jd6

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*, *51*(1), 195–203. 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

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 311–320. https://doi.org/10/dsvgxq

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. [Computer software]. R Foundation for Statistical Computing. 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

Śmigasiewicz, K., Asanowicz, D., Westphal, N., & Verleger, R. (2015). Bias for the left visual field in rapid serial visual presentation: Effects of additional salient cues suggest a critical

role of attention. *Journal of Cognitive Neuroscience*, *27*(2), 266–279. https://doi.org/10/gpckcv

Sherman, M. (2023). Is the N2pc a correlate of attentional selection? An #EEGManyLabs multi-lab registered replication of Eimer (1996). *Peer Community in Registered Reports*. https://rr.peercommunityin.org/PCIRegisteredReports/articles/rec?id=411

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. *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

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10/f7g3rd

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10/fqszj7

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

Wyble, B., Callahan-Flintoft, C., Chen, H., Marinov, T., Sarkar, A., & Bowman, H. (2020). Understanding visual attention with RAGNAROC: A reflexive attention gradient through neural AttRactOr competition. *Psychological Review*, *127*, 1163–1198. https://doi.org/10/gr6gdn

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 *t* test for all pipelines; additional non-parametric test in the bootstrapping pipelines.	We ran a power analysis with 1 - β = 0.90, α = 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 *minus* 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.

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 the Stage 1 Registered Report. We would also like to thank Alexandra Saja, Parmida Mirzaei, Riya

1318 Maliekal, Moon Verma, Eléonore Poinssot-Werlen, Clémentine Marchetti, and all the people
1319 who assisted with data collection in each lab.