This is a repository copy of *#EEGManyLabs : investigating the replicability of influential EEG experiments*.

Article:

Viewpoint

#EEGManyLabs: Investigating the replicability of influential EEG experiments

Yuri G. Pavlov [a,b,*], Nika Adamian [c], Stefan Appelhoff [d],
Mahnaz Arvaneh [e], Christopher S.Y. Benwell [f], Christian Beste [g],
Amy R. Bland [h], Daniel E. Bradford [i], Florian Bublatzky [j], Niko A. Busch [k],
Peter E. Clayson [l], Damian Cruse [m], Artur Czeszumski [n], Anna Dreber [o,p],
Guillaume Dumas [q,r], Benedikt Ehinger [s], Giorgio Ganis [t], Xun He [u],
José A. Hinojosa [v,w], Christoph Huber-Huber [x], Michael Inzlicht [y],
Bradley N. Jack [z], Magnus Johannesson [o], Rhiannon Jones [aa],
Evgenii Kalenkovich [ab], Laura Kaltwasser [ac],
Hamid Karimi-Rouzbahani [ad,ae], Andreas Keil [af], Peter König [n,ag],
Layla Kouara [ah], Louisa Kulke [ai], Cecile D. Ladouceur [aj],
Nicolas Langer [ak,al], Heinrich R. Liesefeld [am,an], David Luque [ao,ap],
Annmarie MacNamara [aq], Liad Mudrik [ar], Muthuraman Muthuraman [as],
Lauren B. Neal [at], Gustav Nilsonne [au,av], Guiomar Niso [aw,ax],
Sebastian Ocklenburg [ay], Robert Oostenveld [x], Cyril R. Pernet [az],
Gilles Pourtois [ba], Manuela Ruzzoli [bb], Sarah M. Sass [bc],
Alexandre Schaefer [bd], Magdalena Senderecka [be], Joel S. Snyder [bf],
Christian K. Tamnes [bg], Emmanuelle Tognoli [bh], Marieke K. van Vugt [bi],
Edelyn Verona [l], Robin Vloeberghs [bj], Dominik Welke [bk], Jan R. Wessel [bl,bm],
Ilya Zakharov [bn] and Faisal Mushtaq [ah,**]

a University of Tuebingen, Germany
b Ural Federal University, Russia
c University of Aberdeen, UK
d Max Planck Institute for Human Development, Berlin, Germany
e University of Sheffield, UK
f University of Dundee, UK
g TU Dresden, Germany
h Manchester Metropolitan University, UK
i University of Miami, USA
j Heidelberg University, Germany
k University of Münster, Germany
l University of South Florida, USA
m University of Birmingham, UK

---

 * Corresponding author. University of Tuebingen, Germany.
 ** Corresponding author.
   E-mail addresses: pavlovug@gmail.com (Y.G. Pavlov), f.mushtaq@leeds.ac.uk (F. Mushtaq).

[n] *University Osnabrück, Germany*

[o] *Stockholm School of Economics, Sweden*

[p] *University of Innsbruck, Austria*

[q] *Université de Montréal, Montréal, Quebec, Canada*

[r] *CHU Sainte-Justine Research Center, Montréal, Quebec, Canada*

[s] *University of Stuttgart, Germany*

[t] *University of Plymouth, UK*

[u] *Bournemouth University, UK*

[v] *Universidad Complutense de Madrid, Spain*

[w] *Universidad Nebrija, Spain*

[x] *Radboud University, Nijmegen, Netherlands*

[y] *University of Toronto, Canada*

[z] *The Australian National University, Canberra, Australia*

[aa] *Department of Psychology, University of Winchester, UK*

[ab] *HSE University, Moscow, Russia*

[ac] *Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Germany*

[ad] *University of Cambridge, UK*

[ae] *Macquarie University, Sydney, Australia*

[af] *University of Florida, USA*

[ag] *University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[ah] *University of Leeds, UK*

[ai] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*

[aj] *University of Pittsburgh, USA*

[ak] *University of Zurich, Switzerland*

[al] *Neuroscience Center Zurich, Switzerland*

[am] *University of Bremen, Germany*

[an] *Ludwig-Maximilians-Universität München, Germany*

[ao] *Universidad Autónoma de Madrid, Spain*

[ap] *Universidad de Málaga, Spain*

[aq] *Texas A&M University, USA*

[ar] *School of Psychological Sciences & Sagol School of Neuroscience, Tel Aviv University, Israel*

[as] *Johannes Gutenberg University, Mainz, Germany*

[at] *University of Texas Permian Basin, USA*

[au] *Karolinska Institutet, Sweden*

[av] *Stockholm University, Sweden*

[aw] *Indiana University, Bloomington, USA*

[ax] *Universidad Politecnica de Madrid and CIBER-BBN, Spain*

[ay] *Institute of Cognitive Neuroscience, Ruhr University Bochum, Germany*

[az] *University of Edinburgh, UK*

[ba] *CAPLAB - Ghent University, Belgium*

[bb] *University of Glasgow, Glasgow, UK*

[bc] *The University of Texas at Tyler, USA*

[bd] *Monash University (Malaysia Campus), Malaysia*

[be] *Institute of Philosophy, Jagiellonian University, Krakow, Poland*

[bf] *Department of Psychology, University of Nevada, Las Vegas, USA*

[bg] *University of Oslo, Oslo, Norway*

[bh] *Florida Atlantic University, USA*

[bi] *University of Groningen, the Netherlands*

[bj] *KU Leuven, Belgium*

[bk] *Max-Planck-Institute for Empirical Aesthetics, Germany*

[bl] *University of Iowa Hospitals and Clinics, Iowa City, USA*

[bm] *University of Iowa, Iowa City, USA*

[bn] *Russian Academy of Education, Russia*

ABSTRACT

There is growing awareness across the neuroscience community that the replicability of findings about the relationship between brain activity and cognitive phenomena can be improved by conducting studies with high statistical power that adhere to well-defined and

standardised analysis pipelines. Inspired by recent efforts from the psychological sciences, and with the desire to examine some of the foundational findings using electroencephalography (EEG), we have launched #EEGManyLabs, a large-scale international collaborative replication effort. Since its discovery in the early 20th century, EEG has had a profound influence on our understanding of human cognition, but there is limited evidence on the replicability of some of the most highly cited discoveries. After a systematic search and selection process, we have identified 27 of the most influential and continually cited studies in the field. We plan to directly test the replicability of key findings from 20 of these studies in teams of at least three independent laboratories. The design and protocol of each replication effort will be submitted as a Registered Report and peer-reviewed prior to data collection. Prediction markets, open to all EEG researchers, will be used as a forecasting tool to examine which findings the community expects to replicate. This project will update our confidence in some of the most influential EEG findings and generate a large open access database that can be used to inform future research practices. Finally, through this international effort, we hope to create a cultural shift towards inclusive, high-powered multi-laboratory collaborations.

1. Introduction

A cornerstone of science is replicability, a fundamental issue that has been at the heart of an intense scientific debate in recent years. An influential report from the Open Science Collaboration (2015), which attempted direct replications of 100 studies from Psychological science in three major journals from the field, indicated that only 36% showed statistically significant findings in the same direction as the original studies, and an average shrinkage of effect sizes by about half. These findings are consistent with a high degree of publication bias (Francis, 2012; Ioannidis, 2005; Kühberger et al., 2014; Sterling, 1959). There are growing concerns that the closely related field of cognitive neuroscience suffers similar issues (Brederoo et al., 2019; Button et al., 2013; Poldrack et al., 2017). Indeed, problems may be even more pronounced in this area, as cognitive neuroscience studies often have small samples and inflated effect sizes (Schäfer & Schwarz, 2019). Further, they are characterised by the use of rich, but also noisy, multi-dimensional data sets, which allows for a multitude of analytical choices (Szucs & Ioannidis, 2017), and thereby the "garden of forking paths" (Gelman & Loken, 2013). Given this context, there is a need to address the replicability of cognitive neuroscience research.

Early work on human electrophysiology presents an interesting anecdote for the value of replication. The recording of electrical oscillations on the surface of a nonhuman primate's cortex was first reported in 1875 (Caton, 1875) and, to the astonishment of the scientific community, in 1929 Hans Berger published the first account of human scalp electrical brain activity (Berger, 1929). From 1929 to 1933, Berger published a series of seminal works showing electrical activity similar (albeit attenuated in comparison) to measures directly from the cortical surface, suggesting that the scalp-recorded signal reflects a genuine activity of human brain function (Davidson et al., 2000). However, the novel signals recorded by Berger showed marked discrepancies with signals recorded from nonhuman animals reported in the literature.

Electrical activity recorded from nonhumans was neither as regular as Berger's demonstrations, nor did it show the 10 Hz signal so prominent in Berger's recording of human participants. Thus, hesitation in believing Berger's findings abounded in the scientific community, and indeed, Berger himself remained somewhat skeptical. Ultimately, a key breakthrough for the use of EEG to study human brain function came in 1934 from Adrian and Matthews (1934; see also Biasiucci et al., 2019) who set out to examine this novel 10 Hz "Berger rhythm". These authors wrote (p. 356):

> *"We found it difficult to accept the view that such uniform activity could occur throughout the brain in a conscious subject, and as this seemed to us to be Berger's conclusion we decided to repeat his experiments. The result has been to satisfy us, after an initial period of hesitation, that potential waves which he describes do arise in the cortex, and to show that they can be explained in a way which does not conflict with the results from animals".*

This independent replication of results was a key contribution to the acceptance of Berger's reports and laid to rest the initial skepticism surrounding the recording of human EEG.

EEG now stands as one of the oldest and the most widely-used investigation techniques in human cognitive neuroscience, with over 6000 publications per year (Pernet et al., 2019, 2020). Yet, while novel EEG findings continue to be generated, replications of such results are scant. The recent fall-out from the Open Science Collaboration has reinvigorated interest in revisiting some landmark studies (e.g., DeLong et al., 2017; Ito et al., 2017; Nieuwland et al., 2018) and inspired a renewed interest in replicating core findings from the cognitive neuroscience literature.

Cognitive neuroscience research is resource-intensive because of equipment cost and complexity, elaborateness of data collection procedures, and computational requirements of data analysis and curation. This often results in studies with small sample sizes and, consequently, with low statistical power. Button et al. (2013) extracted data from 48 meta-

analyses across the neurosciences and estimated the average statistical power to be between ~8% and ~31%. Potential consequences of low statistical power include overestimation of effect sizes, and a reduction in the likelihood that a statistically significant result represents a true effect (Button et al., 2013; Gelman & Carlin, 2014; Vasishth et al., 2018). Ultimately, this produces a situation where results likely have low replicability. A recent examination of 26,841 statistical records reported in 3,801 papers from psychology and cognitive neuroscience indicates that power in cognitive neuroscience is lower than in psychology broadly, with median statistical power to detect small (Cohen's $d = .20$), medium (Cohen's $d = .50$), and large effect sizes (Cohen's $d = .80$) being .12, .44, and .73, respectively. This suggests that the rate of false positives is likely to be in excess of 50% (Szucs & Ioannidis, 2017). A review of 150 randomly selected ERP studies from 2011 to 2017 indicated that the average sample size per group was 21 participants and the statistical power was conservatively estimated as ~.15 for small, ~.50 for medium, and ~.80 for large effect sizes (Clayson et al., 2019). Hence, low statistical power in cognitive neuroscience research casts doubts on the replicability of many research findings.

Another challenge to replicability is known in the literature as "experimenter degrees of freedom" (Simmons et al., 2011). Specifically, analyses can be conducted and the statistics can be computed in many different ways, which allow for "fishing expeditions" to find statistical significance. While these challenges are not specific to cognitive neuroscience nor EEG research, such expeditions are facilitated by the multidimensional nature of neuroimaging data and the multitude of analytical steps involved. For example, in preprocessing signals, a researcher has a high degree of flexibility in decisions about how to deal with artifacts, which filters to apply, and which exclusion criteria to use. Variations in these decisions create opportunities, be it explicit or implicit, to select the processing route that produces the most "preferable" results. A striking demonstration of the impact of analytic flexibility comes from fMRI research, which has similarly multidimensional data as EEG, together with investigator freedom in filtering procedures and other preprocessing steps. When 70 different research teams analyzed the same fMRI dataset with the same hypotheses, they arrived at conclusions that varied dramatically by team (Botvinik-Nezer et al., 2019). For EEG and ERP experiments, it has also been shown that results are sensitive to seemingly subtle differences in preprocessing routines (Robbins et al., 2020). Given this fact, it is surprising that only 63% of data processing pipelines are even reported. The dependence of the results on subtle details of the data processing routines may hinder replication efforts. Furthermore, lack of detail in reporting allows for analytical flexibility to remain hidden (Clayson et al., 2019).

The consequences of analytical freedom in ERP studies were put in the spotlight by Luck and Gaspelin (2017). They presented a detailed analysis of how spurious results can result from choosing specific regions and time windows for analyses based solely on visual inspection of grand average ERP waveforms. This problematic process is referred to as SHARKing, or "Selecting Hypothesized Areas after Results are Known" (Poldrack et al., 2017). Problems are magnified when

results from such practices are presented as hypothesis-driven steps-a process often referred to as HARKing, or "Hypothesizing After the Results are Known" (Kerr, 1998). Other potential degrees of freedom include a number of statistical decisions that can influence the results, such as deciding on the p-value threshold (Benjamin et al., 2018; de Ruiter, 2019; Lakens et al., 2018; see also; Amrhein et al., 2019) or the plausible effect size (Altoè et al., 2020), and choosing between frequentist and Bayesian approaches (van de Schoot et al., 2017). In summary, best practices should limit the possibility to steer results in the desired direction, willfully or not, by post-hoc decisions on data processing, outcome selection, and statistical procedures.

Two options to limit undisclosed degrees of freedom are pre-registration and registered reports. Pre-registration specifies a research plan in advance of undertaking the research and uploading these plans to a publicly available registry. Registered reports are study proposals that are peer-reviewed *before* the research is undertaken. New forms of scholarship and publishing, in which data are shared along with the publication, or directly embedded in manuscripts to allow analysis and re-analysis on the spot (Maciocci et al., 2019) also address some of these issues. It seems inevitable that such approaches will see an increase in popularity in the coming years, but we expect delayed adoption for data-intensive areas of science such as EEG research, due to logistic constraints on voluminous data storage, transfer and online computational power.

Pre-registration and registered reports, coupled with direct replication and systematic documentation of analytical steps, however, remain primary means of assessing the robustness of a given effect (Clayson et al., 2019; Clayson & Miller, 2017; Obels et al., 2020). These same steps, when coupled with larger sample sizes, also allow more stable and precise estimations of effect sizes (Schönbrodt & Perugini, 2013), which are required when translating basic science findings to clinical practice or technological applications. A recent study on the replicability of social-behavioural findings by four coordinated laboratories demonstrated that when original studies and their replications followed methodological transparency and coupled it with higher statistical power and pre-registration, a high rate of replication was achieved (86%; Protzko et al., 2020).

There are a number of barriers towards undertaking replications. Some of these barriers are prevalent across the sciences-it is well-documented that publication pressure tends to incentivise novel effects over incremental research, direct replications (Bradley, 2017) and null findings ("In Praise of Replication Studies and Null Results," 2020). Similarly, research funding bodies have historically prioritised funding for high risk and breakthrough programmes. These issues are compounded by the resource-intensive nature of EEG research. In comparison to most behavioural studies, EEG experiments typically require more resources, such as hardware, taking longer to conduct and analyze. Pooling resources across different laboratories is a potential way to reduce these barriers, but requires establishing shared protocols for equipment preparation and data acquisition, given the potential effects of these variables on ERP phenomena (Melnik et al., 2017).

Over the last decade, major collaborative efforts to increase replicatibility have taken place in the psychological sciences and beyond (Errington et al., 2014; Frank et al., 2017; Klein et al., 2014; Moshontz et al., 2018). As the name of this project ("#EEGManyLabs") reveals, we have been particularly inspired by the "Many Labs" model popularised by Klein et al. (2014), as well as from the examples set by projects such as the Psychological Science Accelerator (Moshontz et al., 2018). This initiative, a large-scale, international replication effort, takes on many replication challenges and aims to test the replicability of some of the most seminal EEG findings. Specifically, we will use a collaborative, multi-site approach and standardized protocol to achieve this aim. In the following sections, we outline our approach, including study selection, sample size determination, and definition of the evaluation process, as well as the expected utility of this project.

2. Project coordination

Given that the burden on any single individual or research group can be high (particularly with the need to collect larger than average samples) while the incentives can be low (e.g., publication biases, lack of funding), this #EEGManyLabs project aims to circumvent barriers to replicating influential EEG studies. Through central coordination and distribution of effort across a large network, we will reduce the resource demands on individual researchers. As illustrated in Fig. 1, to date we have recruited a number of labs distributed across several continents that are willing to participate in this collaborative replication effort.

To overcome many of the administrative issues that come with "big science", we have established an organisational structure (see Fig. 2). The Core Team comprises: (i) Project Coordinators—responsible for general management of the project, oversight and strategic support for all Replication Teams, including planning and establishing communication with and between members of the project; (ii) An Advisory Board of EEG experts——who support the Project Coordinators and provide input on a variety of areas including analyzing EEG, reviewing code, programming of experiments, conducting power analysis, reviewing registered reports, obtaining institutional review board/local ethics committee approvals, applying for funding, and other tasks; (iii) Lead Replicating Labs—individuals or research teams who will take ownership for coordinating a specific target replication. The PI of that lab will be responsible for preparing the registered report for that particular study. In addition to the Lead Replicating Lab, a minimum of two additional Replicating Labs will be included in the Replication Team. The Replicating Labs will be responsible for collecting an agreed upon number of samples and (if possible) analyzing the collected data.

Many of the important decisions made in the creation of this project are described in the following sections and a complete list of all project related decisions and resources is available online (https://osf.io/yb3pq/).

3. Selecting studies for replication

The #EEGManyLabs project aims to assess the replicability of a set of highly influential studies. Given the limited resources and the voluntary nature of the collaboration, we made a pragmatic decision to prioritise investigating highly cited works instead of randomly sampling the literature. Selecting highly cited studies for replication comes with increased interest and motivation from potential replicating labs and followers-key for a community-driven project and consistent with other major replication attempts (Ebersole et al., 2016; Errington et al., 2014; Klein et al., 2018; 2014; 2019).

To identify the most highly cited studies in the EEG literature, we first undertook a systematic search in the Web of Science database, where we extracted the number of citations and normalized by the age of publication (see Fig. 3 and full systematic search protocol at https://osf.io/8qkr3/). To

Number of data collection sites
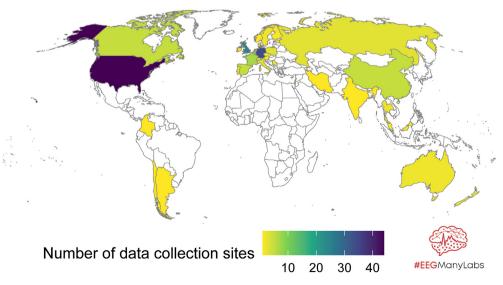
10   20   30   40   #EEGManyLabs

Fig. 1 — #EEGManyLabs Network. Data collection sites include individual researchers or lab groups who have volunteered to collect data for the #EEGManyLabs project. At the time of writing, we have >200 potential data collection sites.
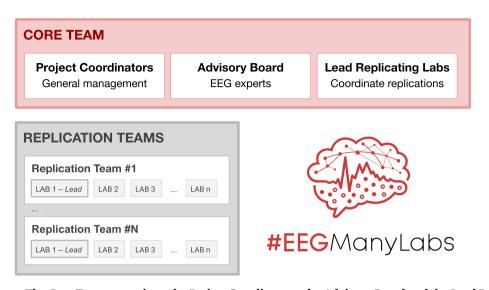
Fig. 2 − **Organogram. The Core Team comprises: the Project Coordinators, the Advisory Board and the Lead Replicating Labs. The Replication Teams are formed for each study by a minimum of three labs.**
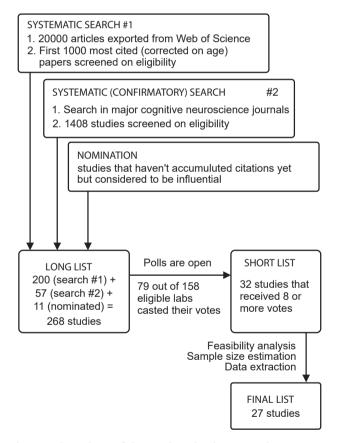

Fig. 3 − **Flow chart of the study selection procedure illustrating how we arrived at the final list of 27 of the most influential EEG studies to be replicated in this project.**

could be conducted in a single session (e.g., longitudinal studies were excluded). Furthermore, we advertised the project on social media (hashtag #EEGManyLabs), inviting the EEG community to nominate studies they deemed worthy of replication. Through social media advertising, we also aimed to identify potentially impactful, recent studies that had not yet had time to accumulate a high number of citations. A more detailed description of the procedure for replication study selection is available online (https://osf.io/8qkr3/). This process resulted in a sample of 268 initial papers for the long list.

To reduce the number of studies considered, the members of the project at the time of study selection (i.e., potential data collection sites—members of the project who expressed willingness to collect data in the future) were asked to cast their votes for the studies they thought to be most influential and worthy of replication. The poll was open to all members, and it was possible for original authors to nominate their own studies. To help researchers identify the studies within their scope of interest, for each of the initially selected papers, a group of volunteers led by the first author (Y.G.P.) manually added keywords describing the main outcome variable (ERP component or EEG measure), studied psychological construct, and other descriptors, including behavioural paradigm used or extra equipment required (e.g., force transducers, eyetracker). This step was deemed necessary because keywords found in the original published papers lacked consistency across studies.

Seventy-nine out of 158 representatives from laboratories expressing a desire to collect replication data at the time of study selection cast their votes. In a third step, the 32 studies that received the highest number of votes (8 or more) were finally selected. The number of votes needed by a study to be selected was arbitrary. It was established to increase the chances of reaching the desired target of at least 20 replications: selecting 41 studies (7 votes or more) would spread the labs thin and selecting 25 studies (9 votes or more) would make the options too scarce. Thus, thirty-two studies entered the feasibility analysis and data extraction stage.

maximise inclusivity and minimise data collection demands, we aimed to include only psychological studies in healthy adult populations using common instrumentation (e.g., no EEG-fMRI), without any special intervention (e.g., no transcranial stimulation, pharmacological manipulation), that

4. Data extraction and sample size estimation

A subset of our team (led by Y.G.P.) was involved in data extraction (e.g., specific hypothesis tests, effect size reached) from the 32 studies selected, to confirm that they all satisfied the minimal criteria for replication. Specifically, we confirmed that (i) each of the key results could be examined through inference tests; (ii) the study employed an experimental or correlational design; (iii) the study examined a topic linking EEG activity and behaviour; and (iv) EEG was used as the primary neuroscience method.

To facilitate replication, the effect of interest needed to be identified and described as precisely as possible in two key ways. First, given that EEG findings are a combination of spatial, frequency, and temporal features, the primary effect of interest needed to be recognized in all relevant dimensions (e.g., "Gamma coherence between visual and somatosensory electrode sites in the 37—43 Hz band was significantly greater during CS + trials than during CS- trials ($p \leq .06$) for the 250-ms time window just before UCS onset"; Miltner et al., 1999). Second, we asked the data extraction team to describe the results in plain-language (e.g., following with the previous example based on the Miltner et al. (1999) study: "Gamma-band coherence increases between regions of the brain involved in an associative-learning procedure in humans").

To determine the upper bound on the sample size required for each replication (the maximum sample size), we extracted the effect sizes from the results reported in the original papers. We assumed the original effect size to be twice as large as it could be in a highly powered study. This assumption is supported by a recent study showing that the effect size in pre-registered studies is about half the size of that in studies without pre-registration (Schäfer & Schwarz, 2019), as well as by the results of large-scale replications (Open Science Collaboration, 2015). To counteract overestimation of the true effects due to publication bias and uncertainty (Brysbaert, 2019), we decided the sample size needed to have 90% power to detect 50% of the original effect size (100% in case of null findings) at a 2% significance level for a one-sided test (see Camerer et al., 2018; Lewis et al., 2020; Schäfer & Schwarz, 2019).

In adopting this approach, studies reporting small effect sizes would require a very large sample size, which could prohibit data collection for many laboratories. At the start of this project, we had asked researchers who were willing to serve as a Replicating Lab how many participants they could contribute to the project. The median number was 50 participants (where range was reported, the maximum number was taken) with only a few labs defining the highest end of the range to be more than 150 participants. Based on this information, we decided to exclude experimental studies that would have required a sample size of more than 200 participants. This led to the exclusion of one experimental study. One further study was excluded because no inference test could confirm or reject the descriptive claim made by the authors. Three studies, focussed on alpha asymmetry, were deemed to be more appropriate as a "spin off" project (see *Legacy*). Following data extraction, 27 potential replication studies remained (Fig. 4).

From a starting position of 27 replication studies, our goal is to conduct replications of at least 20-a number we deem to be a reasonable target that will allow us to generate sufficient data to explore replicability between studies. If we are unable to reach that goal, e.g., due to infeasibility, insufficient number of replicating labs, or rejection at the review stage, we will add the next five studies to the pool from the long list (available at https://osf.io/2qne8/). This procedure will be repeated until the target of 20 replications is met.

5. Prediction markets

Having seen the final list of studies in Fig. 4, many readers familiar with these studies will have their own perspective about the likelihood of individual studies replicating. To what extent they are correct in these beliefs will be the focus of the prediction markets element of this project. Before we collect any data, we will advertise our plans to EEG researchers (including and beyond the #EEGManyLabs network; e.g., social media (#AcademicEEG) and cognitive neuroscience mailing lists) and request their perspectives on the replicability of our target studies in a survey by inviting them to participate in prediction markets. Prediction markets function as a tool to aggregate private information-in this case participating researchers' beliefs about which studies replicate-by giving participants monetary incentives to "bet" on the replication outcomes of the target studies. Previous studies using prediction markets on replications find that they perform better than chance in predicting outcomes and can be considered as an imperfect replication indicator (Camerer et al., 2018; Dreber et al., 2015). We intend to use prediction markets to predict the outcomes of the target replications. At the end of this project, we will be able to examine how closely internally held beliefs in the EEG community map on to the replication results.

6. Modes of participation

There are a number of ways in which individuals and research laboratories can engage with this project. The most critical element of this project is the collection of data. In this section, we detail how we intend to optimise the distribution of data collection across laboratories. Where replications require relatively "large" sample sizes (i.e., >40 participants with analyzable data), a Replicating Lab can decide to collect a smaller sample but distribute the total sample collection among partner labs ("lab buddies") that use the same equipment (with the expectation that, at a minimum, the model of the amplifier and type of electrodes used are identical). Labs with the same amplifier and electrodes will merge their data and form an independent sample to calculate the effect size for internal meta-analysis (the hypothetical study#2 in Fig. 5). For correlational studies, which typically require larger samples, we expect that the distribution of data collection across laboratories will be the default approach. For experimental studies, we require at least three independent samples,
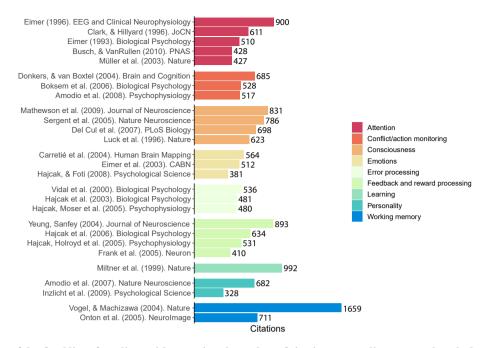
Fig. 4 — **Summary of the final list of studies, with associated number of citations according to Google Scholar as of 01 October 2020. Color indicates the domain of the study. It is important to note that while some studies could have been allocated to multiple domains, we made an arbitrary decision purely for the purpose of visualisation (Amodio et al., 2007a, 2007b; Boksem et al., 2006; Brembs, 2018; Busch & VanRullen, 2010; Carretié et al., 2004; Clark & Hillyard, 1996; Del Cul et al., 2007; Donkers & van Boxtel, 2004; Eimer, 1993, 1996; Eimer et al., 2003; Frank et al., 2005; Hajcak & Foti, 2008; Hajcak et al., 2003, 2005a, 2005b, 2006; Inzlicht et al., 2009; Luck et al., 1996; Mathewson et al., 2009; Müller et al., 2003; Onton et al., 2005; Sergent et al., 2005; Vidal et al., 2000; Vogel & Machizawa, 2004; Yeung, 2004).**

whereas for correlational studies, we require at least two independent samples, with a minimum sample size per replicating lab to be at least equal to the sample size of the original study.

If the required sample size is relatively low ($n \leq 40$), we expect the Replicating Labs to collect the full sample. However, where the sample size expectations are large, it is possible for laboratories to implement a Bayes factors (BF) sequential testing approach (see Schönbrodt & Wagenmakers, 2018), where the target Bayes factor size is specified in the registered reports Stage 1 submission for individual projects, with a maximum sample size and $BF > 6$ recommended to balance feasibility constraints and the level of evidence (Schönbrodt et al., 2017). Once the Bayes Factor indicates sufficient evidence in favor of or against each relevant hypothesis or, alternatively, once the predefined maximum sample size is reached, data collection can be stopped. By offering this flexibility, we aim to minimize any unnecessary use of lab resources and maximize the number of labs willing to contribute.

7. Conducting the replications

Below, we briefly describe the steps each study will go through (see Fig. 6), leaving specific details to the publicly available Project Plan (https://osf.io/yz23p/).

The first step in the replication process is to establish the Replication Team for a particular study. Lead Replicating Laboratories will be self-nominated by filling in a form that will need to be confirmed by the Project Coordinators. After approval, the Lead Replicating Lab will issue a call for Replicating labs, listing all necessary details, such as technical requirements, the expected duration of the experiment, and the planned sample size. After recruiting at least two Replicating labs, the study will proceed to the next stage-development of the study protocol.

The most critical step is to make sure that the replications' methodology closely follows the original and allows the Replication Team to conduct a fair and high-powered test of the main findings from the original study. The Replication Team will prepare the materials (e.g., presentation and analysis scripts) for the replication studies to mirror the methodology used in the original paper. This process will be based on the data extracted from the articles at the stage of selecting experiments for replication and, preferentially, with the original authors' help. The Lead Replicating Lab will have primary responsibility for the development of the new stimulus presentation code in the form of carrying out the task (or identifying suitable people in the Replication Team or wider network who wish to support this activity) and verifying the resulting code. The Replicating Labs will translate the code for stimuli presentation for use in their labs if necessary (e.g., the original is written in E-Prime and provided by the original
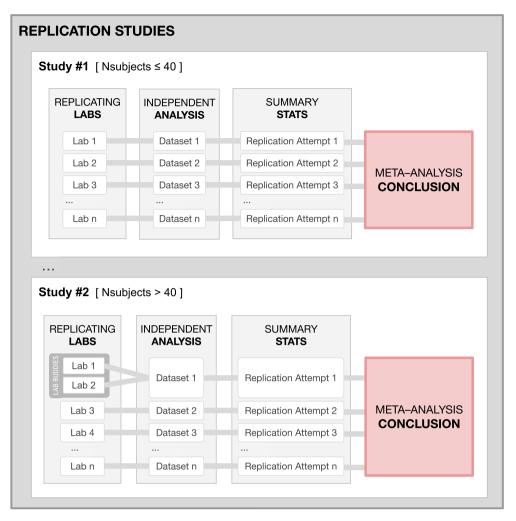
Fig. 5 — Modes of participation for the Replicating Labs. In sample study 1, the agreed-upon number of participants in the study is less than 40, and all labs proceed independently, until the meta-analysis step in which results are combined. In sample study 2, where more than 40 participants are required for each replication study, labs can collaborate and create a joint dataset.

Fig. 6 — A simplified example timeline of a single replication. M indicates month. We expect that there will be considerable variation in timelines for individual replications but that they will follow each of the steps laid out here.

author, but a Replicating Lab uses/has access to Psychtoolbox; Kleiner et al., 2007). When possible, an attempt will be made to write task code using free open source tools (e.g., PsychoPy; Peirce, 2007). The Replication Team will develop the protocol for data collection and analysis based on the available materials and pilot data collected in each of the Replicating Labs. For the sake of transparency and reproducibility, we will give

preference to use open source toolboxes (e.g., Brainstorm (Tadel et al., 2011), EEGlab (Delorme et al., 2011), FieldTrip (Oostenveld et al., 2010), SPM (Litvak et al., 2011)) and free open source software (e.g., MNE Python; Gramfort et al., 2013) in combination with custom-made scripts.

Next, the protocol will be supplemented with an introduction section, including description of the key findings and

a rationale for the original study selection, with clearly stated hypotheses to be tested. The introduction will cover the current evidence for the findings of the original study, paying most attention to any existing studies replicating the original findings, including conceptual replications. The introduction will also stress the impact of the original study and the importance of its replication.

A draft of the manuscript will be reviewed internally by selected members of the Advisory Board for approval. Such a review process is designed to ensure accurate replication of the methods and procedures. Once the manuscript has been internally reviewed, it will be submitted to *Cortex* as a RR Stage 1. Given that a number of notable replications were followed by refutations and criticism from the original authors (e.g., Baumeister & Vohs, 2016; Moran et al., 2020), at this stage in the process, the replicators may wish to have the original authors explicitly endorse interpretations of potential results and confirm the suitability of the planned protocol (Nosek & Errington, 2020). The Lead Replicating Lab can decide to include the original authors as co-authors or to acknowledge their contribution, depending on their level of involvement in the preparation of the RR. To mitigate concerns over the independence of a replication, including biases in the interpretation and discussion of the results, the original authors are allowed to participate only in Stage 1 of the RR.

After in principle acceptance (IPA) and prior to data collection, all methodology, materials, and plans for analysis will be posted in the OSF study registry. The call for Replicating Labs will open up again for research teams who were unable to join the replication team earlier in the development of protocols but have capacity to collect data. Data collection will proceed asynchronously in all Replicating Labs. Replicating Labs will be expected to complete data collection within 1 year of the IPA after obtaining ethics approval from their local ethics committee. If the minimal criterion of having three samples from three independent labs has not been reached, data collection will be extended beyond one year.

The data analysis protocol developed earlier will be used by the Replication Team to analyse the data. All analysis steps will be documented to facilitate re-analysis, and the code will be made publicly available. The analysis scripts in EEG research frequently involve manual artifact identification, correction, and rejection which introduces subjectivity to the process. And while a fully automated preprocessing pipeline has the potential to be more reproducible than one involving manual processing, today's automated algorithms also require some subjective decision making (e.g., defining a numerical threshold for rejection). Given that there is no clear consensus on which approach is superior, we recommend employing the method used in the original study. This should avoid potential non-replications due to deviations in the preprocessing procedure. Where this means manual preprocessing, laboratories will be asked to store trial level data with information on their artefact correction process. In these instances, we also stress that Replication Teams may run supplementary analyses using state-of-the-art automated approaches. In all cases, the teams pledge to abide by a preregistered analysis script. Beyond individual replications, a spin-off team ("#EEGManyLabs Automation") will implement automated analyses to investigate differences between manual and automated coding. Each Lead Replicating Lab will consider whether additional blinding is required during the analysis; e.g., by having manual analyses conducted by researchers who are blind with respect to the experimental conditions. Such blinded analyses will need to be reported as such in the replication attempt. The replicators will be expected to execute the previously agreed analysis script, which will provide an effect size for the meta-analysis. Preprocessed data will be provided to the Lead Replicating Lab for supplementary analyses of the aggregated dataset.

The Lead Replicating Lab will conduct the meta-analysis. The Lead Replicating Lab will report the median and distribution of the weighted and unweighted effect sizes, corresponding 95% confidence intervals, and the number of Replicating Labs successfully replicating the original effect. Effect sizes found by individual Replicating Labs within a Replication Team will be visualized in a forest plot. In addition, the Lead Replicating Lab will delineate the proportion of studies/samples that rejected the null hypothesis in the expected and unexpected direction. Any deviation from the protocol approved at RR Stage 1 will be reported and justified. The contributors of each project will have the opportunity to review and edit the replication manuscript before it is submitted to *Cortex*. Participating labs will also comment on possible explanations for successful/unsuccessful replication.

Replication success is defined operationally as a statistically significant random-effects meta-analytic estimate (at $p < .02$) combining the results from the different laboratories, in the same direction as in the original study. To quantify the variation in effect sizes across samples and settings, the Lead Replicating Lab will further conduct a random-effects meta-analysis and establish heterogeneity estimates to determine if the amount of variability across samples exceeded the amount expected as a result of measurement error.

8.   Data output & management

We intend to share all study materials, complying with *FAIR* principles, making the material *Findable*, *Accessible*, *Interoperable*, and *Reusable* (Wilkinson et al., 2016). The study materials will be shared on OSF (https://osf.io/yb3pq/). All Replication Teams will share their analysis pipelines, preferably in the form of reproducible scripts that include artifact annotations (e.g., visually or automatically identified artifacts, rejected channels, ICA weights, rejected components).

We will inform research participants of the aims of the project, of the experimental procedures, and will explain that research data will be shared. The consent of participants, including to the sharing of their data, is required for their participation. We will use the Open Brain Consent form (Bannier et al., 2020) as a template, which will be adapted to each lab's needs according to their local laws and regulations. Before sharing, raw data will be curated and organized by the Replication Teams following the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016; Pernet et al., 2019), ensuring the removal of any directly identifiable information such as name, address, birth date, etc. By default, minimal demographic data will be requested from each lab (i.e., age,

gender, handedness and education including total years and highest qualification; Pernet et al., 2020) but additional information (e.g., IQ, health or psychological characteristics) might be collected, which will be determined by the Replicating Labs and contingent upon approval from the respective local ethics review boards. Datasets will be shared using a suitable repository (e.g., FigShare, Zenodo, Dataverse) and linked to OSF. We aim to share the data as openly as possible, but depending on requirements imposed by the Replicating Labs local ethics review boards and their institutional and national regulations, the access to shared data may require controlled access, i.e., external interested researchers may have to register and request access. Labs that do not have permission for sharing data cannot participate in data acquisition, but can still contribute to the analysis.

9. Summary report

Once the individual replications of the different studies are completed and published, we will collate and summarise the findings into a summary report, to be published in *Cortex*, that will mark the closing of the direct replication component of the #EEGManyLabs project. This publication will aim to highlight specific and general conclusions from the replicated studies, provide a unified dataset, describe the lessons learned in running this community-driven initiative and ultimately derive recommendations for future EEG research.

While the nature of each replication and their theoretical implications will be dealt with in the individual replication reports, the summary report will focus on aspects that are common across our studies. Our central repository (https://osf.io/yb3pq/) will contain (i) a document summarizing details of the recording setups and data collection according to COBIDAS standard (e.g., the amplifiers used; the number, composition, and layout of sensors; acceptable and observed impedances; recording reference and ground; sampling rates; and acquisition filter bandwidths; see Pernet et al., 2020); (ii) environmental information such as lighting, sound attenuation, and electromagnetic shielding; (iii) pre-registered analysis codes and procedures, accompanied by test data collection videos; and finally (iv) links to all data repositories of the individual replication attempts. Based on this information, we will evaluate how similar the procedures of the replication attempts were to those reported in the original studies (e.g., with regards to sample size; subject and trial-level artifact rejection rates).

Replication outcomes will be summarized with a hierarchical forest plot to illustrate all replication studies' effect sizes. We will also illustrate effect sizes across Replicating Labs for a single study and heterogeneity of effect sizes across labs (i.e., addressing a common "hidden moderators" argument; Bavel et al., 2016). These effect sizes will be directly contrasted with the original papers' effect sizes, and supplemented by reports on *p* value distributions, Bayes factors, and Standardized Measurement Error (SME) measures.

Given the multi-laboratory and multi-experiment nature of this project, we also expect methodological differences across sites and studies to contribute to a proportion of the variance in the results. We will accordingly make a concerted effort to identify the extent to which these factors indeed influence replicability. The impact of these covariates will be examined with respect to the (i) original effect size; (ii) original study design (e.g., within-group *vs* between-group, trial number per condition, sample size, amplifiers used); (iii) data collection parameters (e.g., number of trials, number of channels); (iv) original analysis pipeline/parameters (e.g., reference channel, the complexity of the processing pipeline, how the data were reduced to a univariate inferential test such as averaged quantification across chosen time window and channels *vs* massive univariate testing of all time points and channels with cluster test); and (v) publication characteristics including year of the original studies and journal impact factor, to see whether advances in EEG research practice have improved replicability over the years and whether the profile of the original journal has any relationship with the replicability of a finding. The impact of these factors (and their interactions) will be crucial in recognizing and recommending best practices.

The summary report will also include the outcomes of our prediction markets. The prediction markets will indicate how well researchers in the field can predict the outcomes of the replication studies and whether they under- or overestimate the percentage of studies that replicate. Prediction markets in psychology generally, and neuroimaging (e.g., fMRI) specifically, have been leveraged to provide an index of researchers' ability to judge the replicability of findings in individual subfields (Dreber et al., 2015), but this has yet to be applied to EEG research. By comparing the replicability of EEG studies estimated with prediction markets to actual replicability, we will provide unique commentary on the ability of EEG experts to accurately judge currently published findings as well as the potential of using prediction markets as a future tool to assess the face validity of EEG research.

Finally, we expect to close our summary report with recommendations for further research (both replication and original research), based on the above analyses and the experience gained across the many labs participating in this large-scale project. We expect to identify the minimum number of trials and participants needed to detect some of the most common EEG phenomena (e.g., N2pc, N2 in go/no-go tasks, ERN, P3b) with the help of a sensitivity analysis, and more generally, to make suggestions about recommended parameters in data collection and analysis protocols.

10. Project outcomes

EEG/ERP research on human cognitive processes has been built upon a vast body of data collected over approximately six decades. One of the main strengths of this field is that key effects have been widely replicated (e.g., the P300 responsiveness to infrequent trials in the oddball task), enabling researchers to use ERPs as biomarkers of cognitive processes. However, it is still unclear if many essential findings in this field will withstand the test of direct independent replications, and how much effect sizes differ across laboratories and with larger sample sizes. #EEGManyLabs will help to address these questions by providing a perspective on past work while suggesting tools to improve future research.

This project will provide an initial estimate of the replicability of a set of key findings from studies that were selected by the EEG research community because of their impact on the field. By investigating covariates and moderators of replication successes versus failures, this project can provide knowledge that enhances the replicability of future EEG studies. Outcomes from the replication studies that are consistent with those of the original studies will increase confidence in the original studies' findings and their robustness; conversely, outcomes inconsistent with those of the original studies will decrease confidence in these outcomes and the related conclusions (Nosek & Errington, 2020) and launch a search for explanatory factors contributing to discrepancy between initial and replication studies.

We must also stress the importance of what will not, or cannot, be learned from this exercise. Given the nature of the studies that are to be replicated, it is clear that the conclusions from this project will not apply to all EEG/ERP research. We selected influential (i.e., highly-cited) studies for this project and, as such, this project can only provide an estimate of the replicability of a subset of EEG/ERP research-not the field at large. Indeed, it is possible that the most influential studies might be more or less replicable than studies that have been cited less often. For example, one may make the argument that as highly influential studies often introduce new or exciting findings (i.e., are not incremental), they may be less likely to replicate than studies that advance the field more slowly because they are more closely tied to prior work. Thus, the original Many Labs replications found little difference in replication as a function of citation rate (Altmejd et al., 2019). Another factor to consider here is that our selection process involved a nomination and voting process-perhaps some study selections were based on skepticism. We expect that our prediction markets will uncover these subjective beliefs amongst the EEG community for this set of studies, but alternative approaches will be needed to provide an estimation of the replicability of EEG research more generally.

10.1.   Legacy

Beyond the specific outcomes related to the individual studies, we expect this project to have a long-lasting legacy on EEG research across a broad number of domains. We also hope that this project can provide a canvas for future replication projects of EEG/ERP studies that were not included in the current project. We describe some of the expected legacies of this project next.

As a starting point, we will allow researchers outside the #EEGManyLabs network to access all our replication data and materials to perform future re-analyses in an open and transparent way. We hope that future work will be able to better understand the optimal characteristics of a replicable study. To this end, we will make all the raw and processed EEG replication data available using Brain Imaging Data Structure (BIDS) guidelines, as well as analysis scripts, experimental stimuli, stimuli presentation scripts, lab notes, video recordings, and other research materials.

One longer term benefit of this project will be empirically well-justified recommendations for sample sizes for EEG studies of particular phenomena. Effect sizes will be computed for specific components across a wide range of tasks. Researchers will thus have a database to use when considering how those measures may vary across stimulus characteristics, response demands, trial numbers, and other task parameters. Such data should help inform sample size planning for future EEG/ERP studies.

The #EEGManyLabs project will also result in a series of broader recommendations and practice guidelines on how to conduct multi-site EEG studies in the domain of EEG. The superficially simple task of merging two EEG signals acquired from different amplifiers is far from trivial. By providing data on how variability in the collection of data across sites affect the result, we hope #EEGManyLabs will help future researchers to plan their multi-lab studies and set the scene for future collaborative science.

At the time of writing this manuscript, #EEGManyLabs has already inspired several ongoing and planned projects. One subproject ("spin-off"), #EEGManyLabs Asymmetry, will leverage community engagement to record additional resting-state EEG data, and a set of personality questionnaires together with the replication attempts. In doing so, it will shed light on the replicability of asymmetries in EEG alpha power (Reznik & Allen, 2018) and their relation to personality traits. Another spin-off (#EEGManyLabs Automation) will compare the outcomes of analyses conducted by the #EEGManyLabs Replication Teams with a fully automated analysis pipeline developed by a group of analysts. This project aims to evaluate the within-study effect of manual versus algorithmic artifact removal in the replication context to investigate the role of subjective biases associated with manual coding discussed above. The project will also investigate whether the original studies that implemented automatic artifact rejection algorithms are more often successfully replicated than those that used manual coding methods. In this way, we will be able to address the question of whether automation can help to improve replicability. The datasets generated from this project will also allow us to study the effects of analytical flexibility on EEG findings' robustness in another ongoing project-#EEGManyPipelines. Here, researchers will be invited to analyse the replicated datasets using their preferred analysis pipelines and will then analyze variation in analysis pipelines and the resulting diversity of results.

10.2.   Inclusivity and collaboration

Since the start of the project, we have aimed to establish a wide network of researchers and data collection sites, with diverse scientific interests and skill sets. The current #EEGManyLabs Network represents 33 countries on 4 continents (with hopes to further expand membership-particularly in under-represented countries), and approximately 30% of researchers currently involved are women (identified based on given names using genderize.io database). However, the studies selected for replication all come from Western Europe and North America and are overwhelmingly authored by men.

While the selected studies reflect a broad issue with lack of diversity in research, we are hopeful that the current project will bring much needed diversification to EEG by conducting transparent research, producing open data and materials, and promoting global collaboration.

This brings us to the final goal of this project. Through demonstrating the feasibility of large-scale multi-site projects involving a large, diverse body of EEG researchers, we hope to facilitate a cultural shift away from small-scale single laboratory experiments towards high-powered, community-driven collaborations, creating a stronger foundation for the future of EEG research.

11. Conclusions

In an international effort spanning multiple research institutions and numerous researchers, the #EEGManyLabs initiative promises to yield high-fidelity replication attempts of influential EEG/ERP experiments. Following the Many Labs model (Klein et al., 2014), each experiment will be replicated in several labs to collect a large sample of data for each study, allowing the assessment of replicability through internal meta-analyses. To ensure a high scientific standard is maintained across all replications, this concerted effort is centrally coordinated. Each replication will pass quality control through being reviewed by members of the advisory board, will use standardised experimental and analysis protocols across labs, and involve registered reports that will be published irrespective of the outcomes. A final meta-analytical report will synthesize outcomes from across all replications and will mark the end of this initiative. We expect this project's legacy will rest in pushing the field towards higher replicability standards and facilitating an open science culture of high powered, large-scale multi-site collaborations.

Author contributions

Conceptualization: Yuri G. Pavlov, Niko A. Busch, Damian Cruse, Michael Inzlicht, Andreas Keil, Nicolas Langer, Heinrich R. Liesefeld, Gustav Nilsonne, Sebastian Ocklenburg, Robert Oostenveld and Faisal Mushtaq. Data Curation: Yuri G. Pavlov, Nika Adamian, Artur Czeszumski, Benedikt Ehinger, Xun He, Evgenii Kalenkovich, Layla Kouara and Ilya Zakharov. Methodology: Yuri G. Pavlov, Nika Adamian, Stefan Appelhoff, Daniel E. Bradford, Damian Cruse, Artur Czeszumski, Anna Dreber, Benedikt Ehinger, Michael Inzlicht, Magnus Johannesson, Evgenii Kalenkovich, Andreas Keil, Nicolas Langer, Heinrich R. Liesefeld, Lauren B. Neal, Gustav Nilsonne, Guiomar Niso, Sebastian Ocklenburg, Robert Oostenveld, Cyril R. Pernet, Magdalena Senderecka, Joel S. Snyder and Faisal Mushtaq. Project Administration: Yuri G. Pavlov, Damian Cruse, Gustav Nilsonne, Sebastian Ocklenburg and Faisal Mushtaq. Supervision: Yuri G. Pavlov and Faisal Mushtaq. Validation: Yuri G. Pavlov, Nika Adamian, Artur Czeszumski, Xun He and Evgenii Kalenkovich. Visualization: Yuri G. Pavlov, Guiomar Niso and Magdalena Senderecka. Writing-Original Draft Preparation: Yuri G. Pavlov, Stefan Appelhoff, Niko A. Busch, Damian Cruse, Xun He, Michael Inzlicht, Andreas Keil, Layla Kouara, Nicolas Langer, Heinrich R. Liesefeld, Gustav Nilsonne, Alexandre Schaefer and Faisal Mushtaq. Writing-Review & Editing: Yuri G. Pavlov, Nika Adamian, Stefan Appelhoff, Mahnaz Arvaneh, Christopher S. Y. Benwell, Christian Beste, Amy R. Bland, Daniel E. Bradford, Florian Bublatzky, Peter E. Clayson, Damian Cruse, Artur Czeszumski, Anna Dreber, Guillaume Dumas, Benedikt Ehinger, Ganis Giorgio, José A. Hinojosa, Christoph Huber-Huber, Michael Inzlicht, Bradley N. Jack, Magnus Johannesson, Rhiannon Jones, Evgenii Kalenkovich, Laura Kaltwasser, Hamid Karimi-Rouzbahani, Peter König, Louisa Kulke, Cecile D. Ladouceur, Nicolas Langer, Heinrich R. Liesefeld, David Luque, Annmarie MacNamara, Liad Mudrik, Muthuraman Muthuraman, Lauren B. Neal, Gustav Nilsonne, Guiomar Niso, Sebastian Ocklenburg, Robert Oostenveld, Cyril R. Pernet, Gilles Pourtois, Manuela Ruzzoli, Sarah M. Sass, Alexandre Schaefer, Magdalena Senderecka, Joel S. Snyder, Christian K. Tamnes, Emmanuelle Tognoli, Marieke K. van Vugt, Edelyn Verona, Robin Vloeberghs, Dominik Welke, Jan R. Wessel, Ilya Zakharov and Faisal Mushtaq.

REFERENCES

Adrian, E. D., & Matthews, B. H. C. (1934). The Berger rhythm: Potential changes from the occipital lobes in man. *Brain, 57*(4), 355–385. https://doi.org/10.1093/brain/57.4.355

Altmejd, A., Anna, D., Forsell, E., Huber, J., Imai, T., Johannesson, M., Kirchler, M., Nave, G., & Camerer, C. (2019). Predicting the replicability of social science lab experiments. *Plos One, 14*(12), Article e0225826. https://doi.org/10.1371/journal.pone.0225826

Altoè, G., Bertoldo, G., Zandonella Callegher, C., Toffalini, E., Calcagnì, A., Finos, L., & Pastore, M. (2020). Enhancing statistical inference in psychological research via prospective and retrospective design analysis. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.02893

Amodio, D. M., Jost, J. T., Master, S. L., & Yee, C. M. (2007a). Neurocognitive correlates of liberalism and conservatism. *Nature Neuroscience, 10*(10), 1246–1247. https://doi.org/10.1038/nn1979

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2007b). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology.* https://doi.org/10.1111/j.1469-8986.2007.00609.x, 0(0), 071003012229008-???

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Bannier, E., Barker, G., Borghesani, V., Broeckx, N., Clement, P., Emblem, K. E., Ghosh, S., Glerean, E., Gorgolewski, K. J., Havu, M., Halchenko, Y. O., Herholz, P., Hespel, A., Heunis, S., Hu, Y., Hu, C.-P., Huijser, D., Vayá, M. de la I., Jancalek, R., … Zhu, H. (2020). The Open Brain Consent: Informing research participants and obtaining consent to share brain imaging data. *Human Brain Mapping, 42*(7), 1945–1951. https://doi.org/10.1002/hbm.25351

Baumeister, R. F., & Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science, 11*(4), 574–575. https://doi.org/10/gf5srq.

Bavel, J. J. V., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Reply to Inbar: Contextual sensitivity helps explain the reproducibility gap between social and cognitive

psychology. *Proceedings of the National Academy of Sciences,* 113(34), E4935−E4936. https://doi.org/10.1073/pnas.1609700113

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., … Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), 6−10. https://doi.org/10.1038/s41562-017-0189-z

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten, 87*(1), 527−570. https://doi.org/10.1007/BF01797193

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). *Electroencephalography.* Current Biology: CB, 29(3), R80−R85. https://doi.org/10.1016/j.cub.2018.11.052

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology, 72*(2), 123−132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Botvinik-Nezer, R., Iwanir, R., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., Dreber, A., Camerer, C. F., Poldrack, R. A., & Schonberg, T. (2019). FMRI data of mixed gambles from the neuroimaging analysis replication and prediction study. *Scientific Data, 6*(1), 106. https://doi.org/10.1038/s41597-019-0113-7

Bradley, M. M. (2017). The science pendulum: From programmatic to incremental—and back? *Psychophysiology, 54*(1), 6−11. https://doi.org/10.1111/psyp.12608

Brederoo, S., Nieuwenstein, M., Cornelissen, F., & Lorist, M. (2019). Reproducibility of visual-field asymmetries: Nine replication studies investigating lateralization of visual information processing. *Cortex, 111*, 100−126. https://doi.org/10.1016/j.cortex.2018.10.021

Brembs, B. (2018). Prestigious science journals struggle to reach even average reliability. *Frontiers in Human Neuroscience, 12.* https://doi.org/10/gc5k7j.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*(1), 16. https://doi.org/10.5334/joc.72

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences, 107*(37), 16048−16053. https://doi.org/10.1073/pnas.1004801107

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365−376. https://doi.org/10.1038/nrn3475

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637−644. https://doi.org/10.1038/s41562-018-0399-z

Carretié, L., Hinojosa, J. A., Martín-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: Neural correlates: Automatic attention to emotional stimuli. *Human Brain Mapping, 22*(4), 290−299. https://doi.org/10.1002/hbm.20037

Caton, R. (1875). Electrical currents of the brain. *The Journal of Nervous and Mental Disease, 2*(4), 610.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience, 8*(5), 387−402. https://doi.org/10.1162/jocn.1996.8.5.387

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

Clayson, P. E., & Miller, G. A. (2017). ERP Reliability Analysis (ERA) Toolbox: An open-source toolbox for analyzing the reliability of event-related brain potentials. *International Journal of Psychophysiology, 111*, 68−79. https://doi.org/10.1016/j.ijpsycho.2016.10.012

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

Davidson, R. J., Jackson, D. C., & Larson, C. L. (2000). Human electroencephalography. In *Handbook of psychophysiology* (2nd ed., pp. 27−52). Cambridge University Press.

de Ruiter, J. (2019). Redefine or justify? Comments on the alpha debate. *Psychonomic Bulletin & Review, 26*(2), 430−433. https://doi.org/10.3758/s13423-018-1523-9

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *Plos Biology, 5*(10), Article e260. https://doi.org/10.1371/journal.pbio.0050260

DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: A commentary on Ito, martin, and Nieuwland (2016). *Language, Cognition and Neuroscience, 32*(8), 966−973. https://doi.org/10.1080/23273798.2017.1279339

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., & Makeig, S. (2011). *EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for advanced EEG processing* [Research article]. Computational Intelligence and Neuroscience. Hindawi. https://doi.org/10.1155/2011/130714

Donkers, F. C. L., & van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56*(2), 165−176. https://doi.org/10.1016/j.bandc.2004.04.005

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B. A., & Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences, 112*(50), 15343−15347. https://doi.org/10.1073/pnas.1516179112

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68−82. https://doi.org/10.1016/j.jesp.2015.10.012

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biological Psychology, 35*(2), 123−138. https://doi.org/10.1016/0301-0511(93)90009-W

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology, 99*(3), 225−234. https://doi.org/10.1016/0013-4694(96)95711-9

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

Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *ELife, 3*, Article e04333. https://doi.org/10.7554/eLife.04333

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review, 19*(6), 975—991. https://doi.org/10.3758/s13423-012-0322-y

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy, 22*(4), 421—435. https://doi.org/10.1111/infa.12182

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron, 47*(4), 495—501. https://doi.org/10.1016/j.neuron.2005.06.020

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9*(6), 641—651. https://doi.org/10.1177/1745691614551642

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time* (p. 348). Department of Statistics, Columbia University.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data, 3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

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/10.3389/fnins.2013.00267

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

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005a). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42*(2), 161—170. https://doi.org/10.1111/j.1469-8986.2005.00278.x

Hajcak, G., & McDonald, N., & Simons, R. F. (2003). Anxiety and error-related brain activity. *Biological Psychology, 64*(1—2), 77—90. https://doi.org/10.1016/S0301-0511(03)00103-0

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology, 71*(2), 148—154. https://doi.org/10.1016/j.biopsycho.2005.04.001

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

In praise of replication studies and null results. *Nature, 578*(7796), (2020), 489—490. https://doi.org/10.1038/d41586-020-00530-6

Inzlicht, M., McGregor, I., Hirsh, J. B., & Nash, K. (2009). Neural markers of religious conviction. *Psychological Science, 20*(3), 385—392. https://doi.org/10.1111/j.1467-9280.2009.02305.x

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), Article e124. https://doi.org/10.1371/journal.pmed.0020124

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience, 32*(8), 954—965. https://doi.org/10.1080/23273798.2016.1242761

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196—217. https://doi.org/10.1207/s15327957pspr0203_4

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1—16.

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., Ijzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology, 4*(1), 20. https://doi.org/10.1525/collabra.158

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C. A., Nosek, B. A., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J., Cromar, R., Vidamuerte, D., Gardiner, G., Gosnell, C., Grahe, J. E., Hall, C., Joy-Gaba, J. A., Legg, A. M., Levitan, C., … Ratliff, K. (2019). Many labs 4: Failure to replicate mortality salience effect with and without original author involvement [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/vef2c

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45*(3), 142—152. https://doi.org/10.1027/1864-9335/a000178

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *Plos One, 9*(9), Article e105825. https://doi.org/10.1371/journal.pone.0105825

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S.-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., … Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour, 2*(3), 168—171. https://doi.org/10.1038/s41562-018-0311-x

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

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., & Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience, 2011*, 852961. https://doi.org/10.1155/2011/852961

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

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616—618. https://doi.org/10.1038/383616a0

Maciocci, G., Aufreiter, M., & Bentley, N. (2019). Introducing eLife's first computationally reproducible article. *ELife*. eLife Sciences Publications Limited. https://elifesciences.org/labs/ad58f08d/introducing-elife-s-first-computationally-reproducible-article.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: Prestimulus phase predicts visual awareness. *Journal of Neuroscience, 29*(9), 2725—2732. https://doi.org/10.1523/JNEUROSCI.3963-08.2009

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., & König, P. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data? *Frontiers in*

Human Neuroscience, 11. https://doi.org/10.3389/fnhum.2017.00150

Miltner, W., Braun, C., Arnold, M., Witte, H., & Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associate learning. *Nature, 397*, 434–436. https://doi.org/10.1038/17126

Moran, T., Hughes, S., Hussey, I., Vadillo, M., Olson, M., Aust, F., Bading, K., Balas, R., Benedict, T., Corneille, O., Douglas, S., Ferguson, M., Fritzlen, K., Gast, A., Gawronski, B., Giménez-Fernández, T., Hanusz, K., Heycke, T., Högden, F., & De Houwer, J. (2020). Incidental attitude formation via the surveillance task: A registered replication report of Olson and Fazio (2001). *Psychological Science*.

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature, 424*(6946), 309–312. https://doi.org/10.1038/nature01812

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife, 7*, e33468. https://doi.org/10.7554/eLife.33468

Nosek, B. A., & Errington, T. M. (2020). The best time to argue about what a replication means? Before you do it. *Nature, 583*(7817), 518–520. https://doi.org/10.1038/d41586-020-02142-6

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 229–237. https://doi.org/10.1177/2515245920918872

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *Neuroimage, 27*(2), 341–356. https://doi.org/10.1016/j.neuroimage.2005.04.014

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). *FieldTrip: Open source Software for advanced Analysis of MEG, EEG, and invasive electrophysiological data* [Research article]. *Computational Intelligence and Neuroscience.* https://doi.org/10.1155/2011/156869. Hindawi.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

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

Pernet, C., Garrido, M. I., Gramfort, A., Maurits, N., Michel, C. M., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2020). Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience*, 1–11. https://doi.org/10.1038/s41593-020-00709-0

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B.,

Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience, 18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C. R., Lundmark, S., MacInnis, B., O'Donnell, M., Perfecto, H., Pustejovsky, J. E., Roeder, S. S., Walleczek, J., & Schooler, J. (2020). High replicability of newly-discovered social-behavioral findings is achievable. *PsyArXiv.* https://doi.org/10.31234/osf.io/n2a9x

Reznik, S. J., & Allen, J. J. B. (2018). Frontal asymmetry as a mediator and moderator of emotion: An updated review. *Psychophysiology, 55*(1), Article e12965. https://doi.org/10.1111/psyp.12965

Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., & Bigdely-Shamlo, N. (2020). How sensitive are EEG results to preprocessing methods: A benchmarking study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*(5), 1081–1090. https://doi.org/10.1109/TNSRE.2020.2980223

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

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*(2), 322–339. https://doi.org/10.1037/met0000061

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391–1400. https://doi.org/10.1038/nn1549

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association, 54*, 30–34. https://doi.org/10.2307/2282137

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *Plos Biology, 15*(3). https://doi.org/10.1371/journal.pbio.2000797

Tadel, F., Baillet, S., Mosher, J., Pantazis, D., & Leahy, R. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience, 2011*, 879716. https://doi.org/10.1155/2011/879716

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22*(2), 217–239. https://doi.org/10.1037/met0000100

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language, 103*, 151–175. https://doi.org/10.1016/j.jml.2018.07.004

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

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748—751. https://doi.org/10.1038/nature02447

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Yeung, N. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience, 24*(28), 6258—6264. https://doi.org/10.1523/JNEUROSCI.4537-03.2004