Good Scientific Practice in MEEG Research: Progress and Perspectives

Author names and affiliations

Guiomar Niso[1,2,*], Laurens R. Krol[3,*], Etienne Combrisson[4], Anne-Sophie Dubarry[5], Madison A. Elliott[6,7], Clément François[5], Yseult Héjja-Brichard[8], Sophie K. Herbst[9], Karim Jerbi[10,11], Vanja Kovic[12], Katia Lehongre[13], Steven J. Luck[14], Manuel Mercier[15], John C. Mosher[16], Yuri G. Pavlov[17,18], Aina Puce[19], Antonio Schettino[20,21], Daniele Schön[15], Walter Sinnott-Armstrong[22], Bertille Somon[23], Anđela Šoškić[24,12], Suzy J. Styles[25,26], Roni Tibon[27], Martina G. Vilas[28], Marijn van Vliet[29], Maximilien Chaumon[13,✉]

[1]Psychological & Brain Sciences, Indiana University, Bloomington, IN, USA, [2]Universidad Politecnica de Madrid and CIBER-BBN, Madrid, Spain, [3]Brandenburg University of Technology Cottbus-Senftenberg, [4]Aix-Marseille University, Institut de Neurosciences de la Timone, [5]Laboratoire Parole et Langage, CNRS, Aix-Marseille Université, [6]The University of British Columbia, [7]Google, Inc., [8]Centre d'Ecologie Fonctionnelle et Evolutive, Université Montpellier, CNRS, EPHE, IRD, Montpellier, France, [9]Cognitive Neuroimaging Unit, INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center, Gif/Yvette, France, [10]Cognitive and Computational Neuroscience Lab, Department of Psychology, University of Montreal, Montreal, QC, Canada, [11]Mila - Quebec Artificial Intelligence Institute, [12]University of Belgrade Faculty of Philosophy Laboratory for neurocognition and applied cognition, [13]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France, [14]Center for Mind & Brain, University of California, Davis, CA, USA, [15]Aix Marseille Univ, INSERM, INS, Institut de Neurosciences des Systèmes, Marseille, France, [16]McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA, [17]University of Tuebingen, Germany, [18]Ural Federal University, Yekaterinburg, Russia, [19]Indiana University, [20]Erasmus University Rotterdam, [21]IGDORE, [22]Duke University, [23]ISAE-SUPAERO, Université de Toulouse, France, [24]University of Belgrade Teacher Education Faculty, [25]Psychology, Nanyang Technological University, Singapore, [26]Singapore Institute for Clinical Sciences, A*STAR, Singapore, [27]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK, [28]Ernst Strüngmann Institute for Neuroscience, Frankfurt am Main, Germany, [29]Aalto University, Finland, *authors contributed equally, ✉corresponding author: maximilien.chaumon@icm-institute.org

Highlights

- Good Scientific Practice (GSP) describes recommended methods and procedures.
- GSP minimizes errors and biases, and facilitates collaboration and reproducibility.
- We outline current and developing MEEG GSP insights reflecting the state of the art.
- This overview includes resources, tools, and thoughts to support GSP in MEEG research.
- In all parts of the research cycle, we identify an increased tendency to collaborate.

Abstract

Good Scientific Practice (GSP) refers to both explicit and implicit rules or guidelines that help scientists to produce work that is of the highest quality at any given time, and to efficiently share that work with the community for further scrutiny or utilization.

For experimental research using magneto- and electroencephalography (MEEG), GSP includes specific standards and guidelines for technical competence, which are periodically updated whenever new findings come to light. However, GSP also needs to be periodically revisited in a broader light. At the LiveMEEG 2020 conference, a reflection on GSP was fostered that included explicitly documented guidelines and technical advances, but also emphasised intangible GSP: a general awareness of personal, organisational, and societal realities and how they can influence MEEG research.

This article provides an extensive report on most of the LiveMEEG contributions and new literature, with the additional aim to synthesize ongoing cultural changes in GSP. It first covers GSP with respect to cognitive biases and logical fallacies, pre-registration as a tool to avoid those and other early pitfalls, and a number of resources to enable collaborative and reproducible research as a general approach to minimize misconceptions. Second, GSP with respect to data acquisition, analysis, reporting, and sharing is discussed, including new tools and frameworks to support collaborative work. Finally, GSP is considered in light of ethical implications of MEEG research and the resulting responsibility that scientists have to engage with societal challenges.

Considering among other things the benefits of peer review and open access at all stages, the need to coordinate larger international projects, the complexity of MEEG subject matter, and today's prioritization of fairness, privacy, and the environment, we find that current GSP tends to favour collective and cooperative work, for both scientific and for societal reasons.

Keywords:
Magnetoencephalography (MEG), Electroencephalography (EEG), Good Scientific Practice

1. Introduction

The generation of scientific knowledge, as is the purview of science and the humanities, strongly relies on the honesty and integrity of the scientists and researchers. So too does the public's trust in the scientific community and in its findings. Such integrity and trust must continue to be actively promoted in every discipline, every career, and every single project. To that end, a large number of rules, procedures, and guidelines exist to describe Good Scientific Practice (GSP): ways to practice science so that results are reliable and scientific misconduct is avoided.

In magnetoencephalography (MEG) and electroencephalography (EEG), here collectively referred to as MEEG, the first guidelines for good practice were published for EEG by (Donchin et al., 1977). They represented the findings of an international committee formed to provide publication criteria, which the committee members first discussed among themselves before opening them up for public discussion during a conference. It was here that the recommendation was recorded to, among other things, report age, sex, and handedness of all participants – something still done to this day. Since then, this has been a common pattern: new or additional guidelines have often emerged from international meetings or societies where scientists consider the need for standardization in light of new methods, techniques, or findings.

Currently, the most recent effort in the MEEG community to coordinate and promote GSP at all stages of the research cycle has come from the Organization for Human Brain Mapping. Their COBIDAS MEEG white paper (C. Pernet et al., 2020) lists current standards and good practices for data acquisition, analysis, reporting, and sharing. Together with numerous other guidelines that have emerged over the years (Donchin et al., 1977; Duncan et al., 2009; Gross et al., 2013; Handy, 2005; Hari et al., 2018; Hari & Puce, 2017; Kane et al., 2017; Kappenman & Luck, 2016; Keil et al., 2014; Luck, 2005, 2014; Picton et al., 2000; Pivik et al., 1993), they support scientists and researchers in minimizing known pitfalls and adhering to best practices during the various stages of research.

Such documents are even more important now that the field is interdisciplinary, and not all colleagues will have had the opportunity to be trained in all relevant technologies (C. Pernet et al., 2020). Furthermore, recent concerns about reliability in neuroimaging research (Button et al., 2013; Poldrack et al., 2017) have highlighted the importance of community standards – not just their existence, but also the community's own awareness of, and adherence to,

them. Therefore, neuroimaging in general is actively and effectively working towards consolidating GSP within its community (Poldrack et al., 2020).

Given the current emphasis on GSP, the virtual LiveMEEG conference on "Good Scientific Practices in EEG and MEG Research" was held on October 5-9, 2020, with the aim of bringing together MEEG experts to discuss essential aspects of GSP for the entire lifecycle of MEEG research projects. Importantly, what emerged from this meeting was that not everything that is deemed important is captured by current or previous standards – nor, indeed, could it be: there are many subtleties and approaches that do not lend themselves to being standardized per se. Instead, they represent a more general reflection on the way we think and act, with respect to both the work itself and the broader scientific environment.

Our two-fold aim here is to review existing and developing GSP resources, and to capture those LiveMEEG contributions that widened our understanding of what GSP means. GSP is not merely about adhering to established protocols and avoiding mistakes: it involves a more general awareness of personal, organisational, and societal realities, of the structures that surround and influence us all, and of the future that we ourselves wish to see on the path of scientific progress. This work aims to help foster that awareness.

This paper is organized in three sections. First, we highlight the need for enhanced thinking and preparation, even before experimental data are acquired, the need to take into account the fact that we, humans, are limited in our capabilities, and that there exist procedures and resources to help us expose and mitigate the unintended negative effects of such limitations. Second, we introduce a selection of core GSP in the classical sense of good experimental practice, with recommendations for data collection, analysis, reporting, and sharing. Third and finally, we open up our reflection to encompass the implications of scientific inquiry beyond the laboratory, with thoughts on ethical responsibilities, epistemological stances, and our accountability to the society that supports our work.

2. Thinking about and designing an MEEG project

When planning a project, even before data are collected, some considerations require attention. We begin this section with concerns that affect scientists and researchers themselves: all humans, even trained experts, are susceptible to certain biases that may influence the choices they make. Some biases can be addressed explicitly while designing an experiment. Even so, the process of designing a new experiment involves many details

that can easily be overlooked. Where possible, it is worthwhile to avail oneself of the expertise of others. Therefore, the second subsection deals with ways in which the larger scientific community can be formally involved in experimental design, via pre-registration. Finally, even a flawless experiment is, ultimately, just a single data point: ideally, a true effect is demonstrated multiple times across different settings and contexts. Therefore, even before collecting their own data, scientists should consider how they can support others to collect the same data. This is covered in the final part of this section.

2.1. Human factors

Scientists are trained to form beliefs about the world and test their theoretical and experimental hypotheses with objectivity. However, even expert humans are fallible with regards to reasoning properly, as they can be victims of their own biases, and consequently, tend to make logical errors. Awareness of biases and careful planning of experiments, with attention to the logical implications of given observations, can help mitigate logical errors in science. This section illustrates some of the psychological pitfalls involved with MEEG study planning and execution.

2.1.1. Perceptual biases

The first class of biases we highlight here are called perceptual biases. These often manifest as sensory errors, visual illusions and hallucinations. Because perceptual capacities are limited, humans allocate resources sparingly and create mental shortcuts (heuristics) that are prone to oversights and biases during perception (Gigerenzer, 2008). Evolutionarily speaking, these mental shortcuts are advantageous because they save time and energy. For instance, humans have well-trained mental assumptions about their own size relative to objects in their retinal field; they do not need to measure every doorway before walking through to make sure they fit. But in some less common situations, heuristic assumptions can end up tricking people and give way to illusions (Coren & Girgus, 2020). These perceptual biases can cause individuals to perceive information incorrectly during data analysis and viewing, and end up drawing inaccurate conclusions.

There are many low-level visual illusions that can compromise a researcher’s ability to perceive data visualizations (Coren & Girgus, 2020). For example, color-mapping is a topical issue in the visualization community (Bujack et al., 2018; Munzner, 2014), because human perception of color categories is neither linear nor equal across the color wheel (Bae et al., 2014). This means that using a rainbow color map for continuous data may lead to

perceiving inaccurate steps in the data or meaningless categorical boundaries (Borland & Taylor Ii, 2007) (Cooper et al. this issue). Additionally, although spatial position is precisely represented in human vision (Ware, 2019), viewers of bar and line graphs will typically underestimate average line positions, while they overestimate average bar positions. In cases where the display contains two data series, a “perceptual pull” effect, where the average data positions are “pulled” toward each other, might be observed (Xiong et al., 2020). It is therefore important to keep in mind that a first-glance visual interpretation may be somewhat inaccurate. Although there are not yet empirical guidelines towards mitigating this particular effect, (Xiong et al., 2020) suggest using bars over lines whenever direct positional comparisons are the viewer’s main task (bars showed smaller estimation bias in their study). Another good practice is to present more complete visual representations such as box and density plots (Hintze & Nelson, 1998; Rodu & Kafadar, 2021) that reflect the entire distribution of the underlying data rather than focusing on its central tendencies (Chambers, 2018; Ware, 2019).

Both high- and low-level factors influence the “*curse of knowledge*” effect, where viewers’ own experience and familiarity with a method, theory, or set of results, biases their ability to communicate or predict what others might find salient or important. (Xiong et al., 2020) demonstrated that, once individuals view elements of a line chart as salient or meaningful, they incorrectly predict that new or novice viewers will find the same elements salient or meaningful. Researchers who are planning collaborations should be mindful of this effect when they discuss past work or future study designs. (Xiong et al., 2020) provide three potential mitigation strategies for the curse of knowledge effect in visualizations. The first is to view data from new angles by rearranging values in a chart or using a new encoding type, the second is to critique visualizations with peers, and the third is to rely on the wisdom of the crowd by discussing what others find salient or important and updating personal views. Finally, it is critical that researchers pre-plan their visualization methods so as to avoid the multiple comparison problem in visualization (Zgraggen et al., 2018).

Another way that perceptual bias manifests itself is by way of cultural or academic experience, influencing scientists’ perception of their study results and effects. For instance, (Luo & Zhao, 2019) demonstrated that viewers with different political orientations showed different patterns of visual attention for various elements of global temperature information in climate data visualizations. The high-level experiences of these viewers influenced their low-level perceptual processing of objective information. This same type of bias can influence the reading and understanding of MEEG data, where it is likely that an experimental result which either strongly supports or refutes someone’s own theoretical

predictions will stand out as more salient to them than a less polarizing effect that might warrant deeper consideration. It is therefore critical to think about how individual experiences and expectations might influence the understanding of past work as future studies are planned. Though researchers cannot change their own cultural experience or academic training, they can reflect on their personal belief systems and identify how they might influence the type of information that captures their attention. By acknowledging these perceptual expectations, researchers can be reminded to take a second look at study results and guide their attention to less salient elements in the display. This advice is also practical for the design of experimental stimuli: researchers should consider how their own perceptual biases might be influencing expectations towards participant viewing behavior, especially for emotionally salient displays.

There are many more perceptual biases that can influence the research process (Amer & Ravindran, 2010; Gregory, 1968), but most of them can be addressed by researchers self-reflecting, investigating, and identifying possible distortions, expectations, or assumptions that could be affecting their own work. A first step to avoid them is to be aware of them in the first place.

2.1.2. Cognitive biases

Beyond the sensory nature of perceptual biases, errors can occur at a cognitive level (Baron, 2000). Cognitive biases can also influence data analysis by shaping the top down factors that determine decisions. Much like perceptual biases, there can be much utility found in these heuristics (Gigerenzer, 2008), but researchers should also be wary of their negative effects.

Among the most prominent cognitive biases is a phenomenon called *confirmation bias*, the tendency to select and pay attention to information that confirms our prior beliefs. Combining this bias with the *illusion of validity*, whereby one tends to be overconfident in their own ability to tell a clear story with their data, leads reasoning astray and can create rigid belief systems that hinder scientific progress. Related to this overconfidence is *hindsight bias,* which describes the common feeling that past events were more predictable than they actually were before they occurred. These errors can both amplify the *availability heuristic*, where greater value is placed on information that easily comes to mind, consequently overestimating the likelihood of similar things happening again in the future. In scientific practice, if individuals strongly believe in a particular theory or effect, they may also be strongly resistant to new information which challenges it. A healthy balance should be

achieved, where researchers use critical thinking to evaluate new information fairly and work to either integrate it with their current beliefs or update them as a result of what they have learned should be strived for. This can be accomplished through practice and self-awareness, including thinking of times when researchers' personal or theoretical views were changed with strong experimental evidence, and how they handled the integration of new and old information in that case (see e.g., chapter 4 in Beveridge, 2004).

For data analysis, a frequent reasoning error is *mistaking correlation for causation*. Numerous methods used in MEEG analysis are correlational (including the whole family of methods that range from t-tests and ANOVAs, to generalized mixed models), and "an effect of factor X on the data" should not be taken as evidence for causation (the two could be caused by a third unobserved variable). Instead, specific measures have to be elaborated in order to infer causation. For example, in the realm of connectivity analysis, functional connectivity computed via bivariate correlation between neurological time-courses cannot be taken as evidence that any time-course causes the other one. To reach this type of conclusion in neurophysiological data, directed functional connectivity methods such as Granger causality, which use temporal directionality in the relationships, should be used to infer causality (Seth et al., 2015). Statistical tests and theoretical claims must thus be matched with the methodology and the nature of the phenomenon captured in the experiment.

2.1.3. Social factors influencing reasoning

Social factors influence decision making as researchers conduct and consume science. There are several pitfalls to be aware of. Many researchers feel discouraged (and a few are empowered) by the *hot hand fallacy*, which is the belief that a person who has experienced success has a greater chance of further success in subsequent attempts. While receiving grants and publishing in high-impact journals is clearly not entirely luck-based, there are undeniable elements of serendipity in both processes that must be acknowledged. The inability to challenge these assumptions can lead to individuals following the *bandwagon bias*, simply adopting other prominent ideas and derivative research programs, with the expectation of successfully performing “trendy” research. The *halo effect* is another bias where humans make character judgements based on their overall (often superficial or solely professional) impression of a person, leading to assume for instance that high-impact, tenured research faculty are infallible. When a researcher does find their own success, they must also reflect on the external factors that helped them achieve it to avoid the s*elf-serving*

bias, where individuals take full personal credit for positive achievements and blame external elements, or others, for undesirable outcomes.

Finally, the majority of publications in MEEG research come from Western institutions who recruit *W.E.I.R.D.* (Western, Educated, Industrialized, Rich, and Democratic) participants, meaning that these study results may not generalize to a larger population. In fact, most people in the world are not W.E.I.R.D. (Henrich et al., 2010), which researchers must keep in mind when reading past work and designing new studies, as well as when describing their own experimental conclusions. Lack of equity, diversity and inclusion is unfortunately prevalent in academia, as it is in our countries and cultures. In fact, gender bias, as well as marginalization and underrepresentation of racial, ethnic, and cultural minorities is often enhanced in academia. The MEEG field is no exception to this. It is critical to realize the extent to which women, BIPOC (black, indigenous, and other people of color) and people with disabilities are underrepresented in MEEG research, not only as students and scientists, but also in the participant samples that take part in the conducted experiments. Despite the growing awareness and increasing number of initiatives aiming for diversity and inclusivity, inequalities persist. This impacts negatively not only on the careers, work-life balance,and mental health of underrepresented groups in academia (Llorens et al., 2021; Levitis et al., 2021; Schreiweis et al., 2019), but it also maintains bias and discriminationin our societies. Some relevant resources and concrete actions to mitigate gender bias and inequity at individual and institutional level in academia have been collected by (Llorens et al., 2021). In the field of EEG, recent studies have begun to address the fact that standard EEG electrodes are not designed to accommodate coarse and curly hair common in individuals of African descent (Choy et al., 2021; Etienne et al., 2020). This is an important reminder that systemic bias can also occur through MEEG instrumentation.

The pitfalls of perceptual and cognitive biases can cause problems in research, and should be considered and avoided if possible. However, researchers must also be cognizant of typical errors one can make while interpreting experiments, even in situations not prone to any specific bias described above. The next section introduces an example of how fallacious conclusions can be reached not because of biased reasoning, but because of failure to fully understand and account for the logical implications of an experimental work, and what can be done to avoid misinterpretations.

2.1.4. Logical thinking and fallacious reasoning: an example

Logical thinking is at the heart of conceiving hypotheses and designing experiments. Experimenters need to take certain steps at the outset to forestall later mistakes when drawing conclusions from results.

Some common mistakes can be illustrated by misinterpretations of Libet and collaborators (Libet et al., 1983). In this now infamous EEG experiment, participants were positioned in front of a clock face, and were instructed to freely move their fingers or wrists at a time of their choosing. When they did decide to move, they were asked to report the time when the decision was made. As such, the experiment yielded three points in time: the self-reported time of the conscious will or decision to move (W), the measured time that movement indeed took place (M), and correlated brain activity in the EEG in the form of a readiness potential (RP). They found that the readiness potential RP occurred *before* the reported time of W, which in turn occurred before M. Libet's findings have been replicated often, but there are still raging debates about their interpretation, the role of the neural activity (RP) and W in producing M, and any relation of these findings to free will. Unfortunately, these debates continue to involve several fallacies that could have been avoided by following these simple recommendations.

First, spell out all theories or hypotheses that might explain possible experimental findings. Overlooking an option can lead to a fallacy called *false dichotomy*. For example, commentators on Libet and collaborators (1983) often assumed that one of only three options must be true: a) W causes M, but nothing causes W (Libertarianism); b) W causes an RP, which causes M (Backwards Causation); or c) RP causes M, but W does not cause M (Libet's conclusion). Libet's followers rejected the Libertarian claim, because they assumed that every event has a cause. They also rejected Backwards Causation, because W occurs later than RP, but causes never come after their effects. They concluded that RP causes M, but W does not cause M. What they failed to consider was the Commonsense view that RP causes W, and then W in turn causes M—that is, activity in the brain causes choices, which cause actions. To overlook such a plausible option is called *false dichotomy* because it is false to assume that one member of the original set of alternatives must be true.

Second, think carefully about which theories or hypotheses are logically compatible with others. Forgetting that both of two supposed alternatives might be true can lead to a fallacy called *affirming a disjunct*. Some readers of Libet and collaborators (1983) seem to reason that either RP causes M or W causes M, so, if RP causes M, then W does not cause M (cf.

Sinnott-Armstrong, 2011). However, this reasoning is fallacious because RP and W might both cause M, such as when RP, W, and M form a chain of causes across time.

Third, define the precise category that the conclusion will cover and include stimuli or tasks for all variations within that category. Failing to include a subclass of the phenomenon can lead to a fallacy called *hasty generalization*. An experiment using a specific kind of action cannot justify a conclusion about all actions, especially when the studied acts are atypical. In particular, Libet's subjects arbitrarily chose when to flex their wrists with nothing at stake. What Libet found about these arbitrary, simple, trivial acts might not hold for important human actions that are based on conscious deliberation and require complex sequences. If one wants to draw a conclusion about all actions, one must be sure to include diverse kinds of actions.

How can one know whether one has included enough kinds of test alternatives, considered enough theories, and specified competitors that cannot both be true? The best way to minimize the chances of these and other mistakes is to seek early feedback from others, especially those who disagree with one's views and assumptions. This includes people from other fields and backgrounds. Working together instead of alone is of great benefit: today's team science activities have shown these advantages (see examples in section 2.3. Reproducibility in a collaborative culture).

2.1.5. Example strategy to minimize biases and errors

Experiencing cognitive biases or logical fallacies, and having them influence research is somewhat inevitable, but it is possible to diminish their impact on research practices by learning about them and by using concerted "slow" critical thinking skills (Kahneman, 2013). Table 1: Resources for GSP has a few current references on ways to characterize and circumvent biases and errors. One possible strategy is to organize a "premortem meeting" with fellow researchers before starting a new project (Klein, 2007). This exercise involves simulating a future meeting, after the project *has* failed, and then working backwards from that outcome to determine what could have caused that "failure". This meeting can include a focused conversation about existing biases and beliefs within the group that might influence the work, and then use imaginative exercises to problem solve together. As researchers plan and implement their study, it is important to keep track of which methods are being chosen and why (see also section 2.2. Pre-registration below). At this stage, researchers are encouraged to document which rationales may be biased and how. Skepticism is healthy in science; listening to data is critical, and trying to understand both how data do, or do not, fit

in with existing theories is GSP. A useful set of guidelines come from Abelson's *MAGIC Criteria* (Abelson, 1995), which can be used to understand the *Magnitude, Articulation, Generality, Interestingness, and Credibility* of statistical claims in research. For the many reasons listed in the chapter, once researchers are comfortable operating within a scientific paradigm, it becomes especially challenging to shift their beliefs. But in order to progress science fairly, they must be willing to do so.

- → Perceptual and cognitive biases are consequences of simplification in information processes (e.g. short-cuts or rules of thumb), and pertain to most brain functions.
- → Perceptual biases can affect the way data are perceived, while cognitive biases can affect our strategic understanding of information, which can in turn misguide experimental planning and conclusions.
- → While biases are unavoidable, it is possible to reduce their effects on research practice by learning to characterize and consider them ahead of time.
- → Training, and careful reasoning, laying out all possible outcomes of an experiment is also crucial in that it prevents common logical errors.
- → Collective exercises such as a "premortem meeting" can help highlight pitfalls, and correct mistakes before they are made during an experimental project.

2.2. Pre-registration

One successful practice to mitigate perceptual and cognitive biases and increase procedural transparency is to "pre-register" scientific projects, i.e., define and disclose experimental plans before data are collected and/or analyzed.

To analyze data as rich and complex as MEEG, analysis pipelines combine a multitude of pre-processing and analytical steps, including identification of malfunctioning sensors, data filtering, artifact removal, time-frequency transforms, and so on. Each step involves setting numerous parameters to specific values, driven in part by subsequent analysis requirements, resulting in an extremely large number of combinations to choose from, a problem known in the literature as the "garden of forking paths" (Gelman & Loken, 2013). Given that it is difficult to test the independent contribution of each combination (though see Hoffmann et al., 2020, and Steegen et al., 2016, for a proposed approach), novice and even expert neuroscientists are often left with not knowing how strongly a given effect (or the absence thereof) depends on their analysis choices. Without detailed planning in advance, researchers often make data-dependent choices while analysing data, which inherently

undermines the logic of statistical tests designed for a priori set hypotheses, even if multiple comparisons are not explicitly conducted (Gelman & Loken, 2013; Luck & Gaspelin, 2017). Worse still, they might be tempted to choose the path that leads to “success” (i.e., positive results matching their hypotheses, a pitfall also called "p-hacking"; (Simonsohn et al., 2014)).

Unexpected, ambiguous, or “negative” results are an essential part of scientific discovery. However, the academic incentive system rewards conclusive results and concise stories with publications in “high-impact” journals. Ambiguous, non-significant, or null results often are not publishable, resulting in the infamous “file drawer” problem (Rosenthal, 1979) and the good work of the researcher remaining unrecognized. This mismatch between the academic incentive system and the reality of scientific work ("getting it published" rather than "getting it right", (Nosek et al., 2012)) leads to publication bias and unreliable practices in published work. This is reflected, for example, in a peculiar prevalence of *p*-values just below the critical (yet arbitrary) cut-off of 0.05 (Masicampo & Lalande, 2012) and inflated effect sizes in published studies (Button et al., 2013).

To mitigate publication bias, selective reporting, and suboptimal research practices, the scientific community needs to redefine their incentive systems and develop a culture in which good, transparent scientific practices align with measures of academic success. Pre-registration is one option for increasing the transparency of the scientific process.

The term “pre-registration” encompasses different procedures that occur at different stages in a project’s life cycle, with different levels of formalization. Below, we distinguish between “unlocked” pre-registrations during the project planning phase, and two types of “locked” pre-registrations, namely locked pre-registrations and Registered Reports. According to the project aims, timeline, and career goals of the researcher, certain types of pre-registration might be more optimal.

2.2.1. Unlocked pre-registration

Unlocked pre-registration involves projects at a preliminary stage of planning. At this preliminary stage, the project’s experimental design and analytical approaches are openly discussed, reviewed, and adjusted in the general scientific arenas. These aspects of the project remain “unlocked” and are subject to changes following peer feedback. Some online pre-registration platforms adhere to this preliminary stage and allow revision of the initial protocol (see Table 1: Resources for GSP). A recently suggested alternative for this purpose is that of “prereg posters”, i.e., conference posters that present planned scientific projects

(Brouwers et al., 2020; Tibon et al., 2018). Irrespective of the chosen platform, spelling out the hypotheses of the study as well as criteria for data curation and analysis pipelines before data collection is the common feature of the unlocked pre-registration process, which can take place internally or publicly. Regardless of the platform used, the presentation of research plans prior to data collection allows researchers to receive feedback on their hypotheses, design, and analyses from colleagues, a process that will likely improve the study. In turn, this can facilitate more formal (locked) pre-registration and/or submission of the work as a Registered Report.

2.2.2. Locked pre-registration

Once the protocol is finalized, it is amenable to a formal or "locked" pre-registration, a time-stamped protocol on a public website such as OSF Registries or ClinicalTrials.gov which describes the experimental and analytical procedures of the study in detail (see Table 1: Resources for GSP). The locked protocol is followed throughout the execution of the study, with any deviations clearly reported and justified in the final research paper. This type of pre-registration can help emphasize the planning phase of a study (also as an educational resource for trainees), increase procedural transparency, and serve to publicly take credit for an ongoing study or hypotheses/theoretical models.

2.2.3. Registered Reports

A study protocol can also be locked via the submission of a "Registered Report". This is a published article format that was pioneered by the journal *Cortex* in 2013 (Chambers, 2013) and has since been implemented by more than 250 journals (Chambers & Tzavella, 2020; Nosek et al., 2018). Registered Reports require the submission of a study protocol: introduction, methods, and hypotheses, including criteria for data rejection and stopping rules for testing. The protocol undergoes editorial and external peer-reviews, possibly several rounds of revision, and in the best case ends with "in principle acceptance" (stage 1), where the journal commits to publishing the final report irrespective of whether the results match the initial hypotheses. In a second phase following data collection and analysis (stage 2), the full study is submitted and reviewed again, with guaranteed publication as long as the approved protocol was followed (with changes sufficiently documented and justified).

The clear benefits of Registered Reports for individual researchers are: 1) Motivating the researcher to thoroughly spell out the design, analyses, and hypotheses in a written form with emphasis on the planning phase of the study; 2) Expert feedback on the research plan

before data collection; 3) Acceptance for publication, independent of statistically significant results; and 4) Speeding up final publication by trading off time between initial submission (stage 1) and final publication (stage 2).

For the scientific community in general, benefits are: 1) Increased transparency of experimental procedures, likely enhancing verifiability and reproducibility; 2) A reduction of suboptimal scientific practices (although the impact of pre-registration alone on such practices is debated (Devezer et al., 2020; Rubin, 2020)); and 3) Reduction of publication bias in the scientific literature. Interestingly, with respect to the last point, a comparison of registered versus conventional reports revealed a striking imbalance of 44% versus 96% of positive findings, respectively (Scheel et al., 2021).

2.2.4. Potential concerns regarding pre-registration

Despite the prominent benefits of the pre-registration approaches outlined above, there are some commonly raised concerns. First, “publicly pre-registered protocols might be scooped”. However, with locked pre-registrations and Registered Reports, the time-stamped official report provides proof of when the study was originated by the researcher, and should therefore alleviate these concerns. This could be particularly beneficial for research teams working in competitive fields or with fewer resources.

Second, “the project’s progress incurs a delay”. Namely, for Registered Reports, the review process for stage 1 might add months before data collection can start, which might not be feasible for short project durations or for projects investigating an acute phenomenon. If time permits, however, this can be turned into a benefit, as it shifts some of the heavy load from the publication phase to the planning phase and results in more rigorous protocols prior to data collection. Furthermore, as both the authors and the reviewers have agreed upon an analytical pathway, the temptation to “try alternative pathways” is greatly reduced, and this can be a substantial time saver during the processing and analysis stage.

Third, “proficiency is required for all types of pre-registration”. Coming up with a detailed analysis protocol, especially for complex methodologies like MEEG, requires considerable experience in making informed decisions and utilizing existing pipelines (perhaps from other laboratories) (Paul et al., 2021). Nevertheless, pre-registration can still benefit trainees and inexperienced researchers, with support from their mentors and supervisors. For instance, they can gain the necessary expertise by starting with a replication of a previously published study, where experimental and analysis parameters are more easily derived. Moreover, the

complete preprocessing pipeline (and some of the analyses) can be performed on pilot data, obtained prior to pre-registration. Standardized study pipelines (see Table 2: MEEG analysis toolboxes), templates (e.g., COBIDAS MEEG), and guidelines (e.g., (Paul et al., 2021)) are also excellent means to narrow down the maze of possibilities in MEEG analysis (a few more examples are also discussed in section 3.2 on Good analysis and reporting practice). Thus, a research lifecycle that includes pre-registration provides methodological and research training opportunities, even prior to the pre-registration step.

Finally, "such a strict format would hinder the creative process of science". It is important to note that locked pre-registrations and Registered Reports *do not prevent exploratory analyses or reporting of unexpected findings*. The only requirement is that any addition/change to the original protocol is adequately labeled as such, so as to keep track of the original idea and analysis plan in the published report.

Despite their clear merits, pre-registration and Registered Reports are not the only answer to the replication crisis, nor do they guarantee scientific integrity. However, together with other good practices such as standardized pipelines and open data and code, they are an important tool for improving the transparency and credibility of published science.

- → Pre-registration is the practice of publicly disclosing and archiving experimental plans before data are collected and/or analysed.
- → Pre-registration procedures can be broadly classified as either "unlocked" (publicly available research plans that might still undergo modifications) or "locked" (finalized research plans, which should be followed precisely throughout the execution of the project).
- → Pre-registration has been argued to lead to better scientific practices through increased transparency of the scientific process, peer-review at early stages of the project, and reduced publication bias.
- → Potential concerns or limitations such as "scooping", increased project durations, or reduced applicability for trainees, should be carefully considered but are often mitigated when planning is done carefully.

2.3. Reproducibility in a collaborative culture: some examples

The recent reproducibility crisis in neuroimaging (Button et al., 2013; Poldrack et al., 2017) has highlighted the importance of conducting reproducible research (Poldrack et al., 2020) and stressed its critical value in GSP. While reproducibility is often used as an umbrella term in the literature, here we would like to provide more specific definitions for each of the

dimensions of the process: reproducibility, replicability, robustness, and generalizability, as introduced by (Community et al., 2019). First, scientific work can be defined as *reproducible* when the same results can be systematically observed if the same analytic steps are carried out on the same dataset as the original study. Research is considered *replicable* if the same analysis performed on a different dataset (e.g., another group of participants) outputs qualitatively similar results. *Robust* scientific work occurs when the same dataset is used with a different analysis pipeline that nonetheless tackles the same research question as the original work (e.g., the original code in Python is translated to R, or when the same pipeline is run using different analysis toolboxes) and similar results are obtained. Finally, *generalizable* work combines new but similar datasets, with new but similar analysis pipelines, to produce similar results. Reproducible, replicable, robust, and generalizable research is essential to make scientific progress and thus critical for GSP.

In this section we highlight four initiatives that deal with reproducibility and replicability in an open collaborative culture. First, we introduce the Open Science Framework as a general platform that supports collaboration and reproducibility. For those using such platforms or otherwise interested in conducting collaborative and reproducible research, The Turing Way is discussed next, providing instructions and guidelines on a variety of relevant topics. Following that, we describe #EEGManyLabs, a broad project that uses collaborative practices to replicate a number of influential experiments. Finally, as an example of one project that has already successfully replicated influential ERP experiments and made all material freely available, we discuss the open resource ERP CORE.

2.3.1. Open science framework (OSF)

Conducting reproducible research often involves engaging in efficient collaborative work and coordinating different actors across a wide range of activities. Consequently, researchers are turning to comprehensive services that facilitate collaborative workflow with colleagues and evaluators. One of the most popular is the Open Science Framework (OSF; *https://osf.io/*), a free online platform developed by the non-profit organization Center for Open Science. The OSF allows researchers to manage, document, and share all the products of their workflow, from the pre-registration of the initial idea to the preprint of the final report. This and other resources and general purpose tools are listed in Table 1: Resources for GSP.

2.3.2. The Turing Way

Openly available instructional resources on how to conduct reproducible and replicable work are needed so that every researcher, regardless of background and location, can access knowledge on applying open science practices for collaborative projects, and make proper use of the tools that are available.

The Turing Way (TTW, (Community et al., 2019)) is an example of an educational tool for GSP (see Table 1: Resources for GSP). TTW is an open source, community-developed online collection of guides to reproducible, ethical, inclusive, and collaborative data science. Here we highlight: 1) *Guide for Reproducible Research*: The online book's first guide provides general recommendations for implementing reproducible research practices, such as making sure the computational environment of the project is stored and shared, using a version control system, testing the code, etc. 2) *Guide for Project Design:* This guide covers how to effectively plan and design a research project. Among others, some of its sections illustrate how to write a detailed project roadmap that gives an overview of the goals and the steps to achieve them, how to structure a project's repository (e.g., folder organisation), and how to clearly set-up contribution pathways when working openly. 3) *Guide for Communication:* This guide offers recommendations for communicating research to wider audiences (e.g., through blog posts, podcasts, or social media), or to a specialized public (e.g., during posters and conferences talks). Special emphasis is made in making sure that the communication is accessible to everyone. 4) *Guide for Collaboration:* This guide gives advice for collaborative projects. It includes resources for setting-up a remote collaboration, suggestions for improving the diversity of the team, and an overview of the use of GitHub for open source projects. 5) *Guide for Ethical Research:* This guide discusses the main concepts and institutions associated with ethical research, and provides examples of them through real-world case studies. Finally, TTW has an extended and detailed Code of Conduct that can be used as an inspiration for new collaborative projects. Having a clear and well defined Code of Conduct as part of any collaborative project is of utmost importance to prevent human conflicts overshadowing collective endeavours. This Code of Conduct explains how contributors are expected to behave when collaborating (and what to do if a violation of these expectations occurs), critical for ensuring a welcoming and inclusive collaborative space. TTW is a living documentation open for contributions to accelerate reproducibility practices and GSP.

2.3.3. #EEGManyLabs

Concerns regarding the replicability of psychological phenomena have been spreading to multiple subfields of psychological science (Collaboration, 2015). The scale of the replication crisis in MEEG research has yet to be defined but, given the relatively noisy data, challenges to acquire data from large numbers of participants, and the analytical flexibility in MEEG analysis, MEEG research is unlikely to be free from replicability concerns. In response to this, the ongoing #EEGManyLabs project (Pavlov et al., 2021) aims to revisit a number of the most influential psychophysiological EEG findings with largely increased statistical power to provide a first estimate of the base rate of replicability in EEG. To reach the goal of replicating at least 20 studies in at least three laboratories each, 27 studies have been selected. To ensure high quality, each replication effort takes advantage of the Registered Reports format (see section 2.2.3. Registered Reports), uses standardized operating procedures and analysis pipelines, and passes internal review by the advisory board and original authors.

This extensive commitment to collaborative and open research is expected to increase collective confidence in EEG research, inspire new standards for reporting EEG findings, and provide researchers with a large open database for further exploration. The data collected by #EEGManyLabs will also be combined in an effect sizes catalogue of various commonly studied EEG phenomena, including ERPs, in order to support the design and initiation of novel research. The catalogue addresses one of the most long-standing methodological issues in neuroscience: the problem of deciding the appropriate sample size, for which most methods typically used for *a priori* estimation (e.g., using pilot data, meta-analyses, and single studies) have limitations and are susceptible to bias (Lakens, 2021).

Besides the direct outcomes, the project aims to facilitate a shift towards collaborative psychological science and neuroscience. Studying increasingly subtle EEG effects requires large sample sizes, which are generally not easily achievable in a single lab. For EEG research to continue to thrive in the 21st century, multi-site high-powered studies are required, improving statistical power, reproducibility/replicability, and generalizability. The network of researchers and infrastructure built for #EEGManyLabs is opening new opportunities for testing novel hypotheses in the same collaborative manner in future studies. Moreover, the necessary close collaboration within replication teams should facilitate exchange, spread, and easier adoption of GSP in individual labs. Everyone is welcome to join the network using the links on the project OSF page (see Table 1: Resources for GSP).

In addition to #EEGManyLabs, similar collaborative efforts have emerged in the past several years, for example, multi-site replication initiatives by (Nieuwland et al., 2018), (Nave et al., 2020), and (Whiteford et al., 2020), or a recent ManyLabs initiative to investigate effects of variability in the pre-processing and analysis pipeline (EEGManyPipelines). So far, they are still only sporadic, but may be the beginning of a more widespread trend. These initiatives and resources are listed in the Table 1: Resources for GSP.

2.3.4. ERP CORE

The replicability of MEEG research can be increased by taking advantage of information about optimal design and analysis provided by prior studies. However, ground truth is not typically known, making it difficult to know if the methods from a given prior study are actually optimal and generalizable. The ERP CORE (Compendium of Open Resources and Experiments; (Kappenman et al., 2021)) (see Table 1: Resources for GSP) was created to provide a reference point for future research by taking widely replicated MEEG paradigms, optimizing them, and providing information about them that can be used as solid ground for subsequent studies.

This open resource (Kappenman et al., 2020) contains stimulus presentation scripts, data from 40 neurotypical individuals, and data analysis scripts for six common MEEG paradigms that are designed to isolate seven common ERP components: N170, mismatch negativity, N2pc, N400, P3b, lateralized readiness potential (LRP), and error-related negativity (ERN). These effects have been replicated so many times that the question is not whether they exist, but rather how best to obtain them. The specific versions of the paradigms in the ERP CORE were developed in collaboration with multiple experts with the goal of creating optimized versions that produce valid and reliable effects with approximately 10 minutes of data collection per task. Years of development time was spent on refining the corresponding data analysis pipelines. Moreover, the online resource includes extensive information about data quality and effect sizes that can be used as a reference for comparison with new studies and to conduct power analyses. Finally, data-driven recommendations have been provided for time windows and electrode sites that are optimal for quantifying amplitudes of each component, which can be used to provide *a priori* analysis parameters for future research. This may be particularly useful for investigators acquiring MEEG data with low density/portable systems.

The individual ERP CORE tasks can simply be inserted without change into other studies (e.g., large-scale clinical studies that include an ERP measure) or used as a starting point for new tasks. The data processing pipelines can be used without change to provide *a priori* analysis methods or used as a starting point for new analyses. The relatively large existing dataset can be used to test new hypotheses (e.g., regarding correlations among components) or to assess how new analysis methods work across a variety of paradigms.

→ Reproducible, replicable, robust, and generalizable research is essential to make scientific progress.

→ Tools exist to help prepare experiments in such a way to emphasize reproducibility, replicability, and robustness.

→ The Turing Way is a community-driven guide to reproducible, ethical, inclusive and collaborative data science.

→ #EEGManyLabs is a large-scale Initiative for the replication of some of the most influential EEG papers.

→ ERP CORE is an open resource to reproduce and replicate common ERP experiments.

3. Collecting, analyzing, reporting, and sharing

After carefully planning an experiment, the next steps involve acquiring data, analysing it, and reporting on the results. We first introduce some practical aspects of properly collecting MEEG and its variants (e.g., intracranial EEG), their combination, or the use of simulations where appropriate. We stress the need for having rigorous, documented protocols, and specific points to pay attention to in clinical settings, or with multimodal data. We then highlight GSP for data analysis, with considerations about toolbox usage, signal processing, and statistics, before turning to advanced analysis, data organisation, and reporting frameworks and reporting schemes that ensure maximal reliability and replicability of data analysis.

3.1. Data Collection

Data collection is often the next step after the project has been properly planned. Below, we describe common acquisition protocols for scientific studies, present some resources for multimodal data collection, and introduce some relevant considerations for research studies

in clinical settings. Of course, not every scientific project requires acquiring new data. Open repositories (where pre-recorded datasets can be downloaded free of charge) or carefully designed simulations can allow some projects to go without recording any data at all. Therefore, we will also present some examples of MEEG open data repositories and tools for data simulations.

3.1.1. Common acquisition protocols for scientific studies

It is something of a miracle that we can place an electrode — a small metal disk — on the scalp and record electrical potentials that are generated inside the brain. It is even more miraculous that we can record the tiny magnetic fields that accompany the electrical potentials. The recordings contain brain activity, but these signals are tiny and are mixed with a variety of biological and nonbiological noise sources that can dwarf the neural signals. Substantial effort and ingenuity are required to extract reliable and meaningful brain signals from the noise. Traditionally, some sources of noise in EEG recordings were reduced by abrading the top layer of the scalp and adding conductive gel or solution to improve the low-impedance contact between the electrode and the living tissue of the skin. Today many researchers use high-impedance systems for reasons of speed and safety. Unfortunately, these systems are prone to an increase in skin potentials and other sources of low-frequency noise, especially when the recording environment is warm, which can in turn reduce the statistical power of the study (Kappenman & Luck, 2010). Another important technical factor is an impedance mismatch between an active EEG recording sensor and the reference electrode. This mismatch will undermine the common mode rejection of the EEG amplifier and result in an increased amplification of non-EEG signals. Many other factors also impact the noise level, such as facial, head and body movements, muscle tension, perspiration, and nearby electronic devices. Thus, we are trying to record a needle of brain activity in a haystack of noise. To make matters worse, the most straightforward signal processing method for reducing noise — filtering — can cause significant distortion of the signals in the time domain, shifting latencies and even producing artificial peaks and oscillations (Tanner et al., 2015; Widmann et al., 2015; Yael et al., 2018).

Despite the challenge of dealing with noise, most efforts at developing optimal EEG recording methods (beyond the hardware, e.g., ‘active recording’ systems) have occurred in the clinical setting. Promoting optimal methods so that all research labs are recording the cleanest possible data is crucial (Sinha et al., 2016). Two steps have recently been taken in this direction.

First, a group of researchers has published a detailed EEG recording protocol that provides a precise description of the “special sauce” that they developed over decades for maximizing the signal-to-noise ratio (SNR) (Farrens et al., 2020). For example, participants are asked to vigorously comb their scalps prior to electrode application, and the electrode wires are oriented during electrode application to minimize lead tension and movement artifacts. Details like these do not typically appear in Methods sections of journal articles. A revised version of the protocol is also available with modifications designed for safe testing during the COVID-19 pandemic (Simmons & Luck, 2020). In addition to sharing these tried and tested methods, this protocol also serves as a model for other researchers who wish to publish their protocols. It is published on Protocol Exchange (see Table 1: Resources for GSP) — a free protocol repository that also provides an automatic way to create a digital object identifier (DOI), and indexing on Google Scholar. Other researchers are strongly encouraged to publish their recording protocols, which should increase transparency and reproducibility as well as spreading more effective recording methods.

A second step moving the field toward more optimized recording methods was to develop a new metric of data quality for averaged ERPs, named the standardized measurement error (SME) (Luck et al., 2021). If researchers were to regularly compute the SME and report it in publications, the field could accumulate objective information about which recording methods are best for a given paradigm or measure. The SME is described in more detail in section 3.4. Statistics.

3.1.2. Multimodal data collection: a software solution

MEEG experiments are inherently multimodal. Aside from the MEG or EEG itself, the event markers that indicate the experimental conditions and procedures are recorded simultaneously (if these are necessary). In many cases, additional modalities are also recorded, such as button presses, audio or video, an electrocardiogram, electrooculogram or other signals from the body or peripheral nervous system, eye tracking, etc. This can result in complex set-ups where each modality is running on a different device, at different sampling rates, relying on a different clock to provide time stamps, and using different data formats. The clock issue in particular is critical: when the data are not properly synchronized, it can be impossible to perform accurate analyses.

To solve this synchronization issue, experimenters often employ hardware solutions that connect the different modalities to each other. For example, digital pulses representing experimental events can be sent through a parallel port to the MEEG amplifier,

synchronising the two modalities. However, devising hardware solutions to synchronize all modalities can be cumbersome and impractical, especially when additional hardware has to be added only to synchronize two other pieces of hardware.

Instead of hardware, experimenters can use an increasingly popular software solution called the lab streaming layer (LSL) (Stenner et al., 2021) (see Table 1: Resources for GSP). This is a free and open source software framework that allows for the transmission and collection of data across one or more devices on a local network. When transmitting and recording data through LSL, the software automatically takes care of data formatting, networking, and synchronization. Many manufacturers of experimental hardware, be it MEEG, eye tracking, virtual reality, or motion capture, already support LSL, and a collection of additional apps exist to record many different modalities. For programmers, there is a core LSL library written in C++ with bindings for many other languages, including Java, MATLAB, and Python.

There are a number of other advantages to using LSL, such as its automatic support for data buffering and reconnection in case a network connection is temporarily lost, its support of metadata, and its open file format in which all modalities can be recorded simultaneously. Still, LSL does not solve all experimental issues. In particular, while it can synchronize two pieces of hardware, it does not know the time-stamp of the sample as it is provided by the hardware, or the delay between the software and the hardware. Therefore, a hardware-based measurement of a sample's or device's delay may still be required. However, since this delay should be constant, this measurement is needed only once, and can be taken independently of any particular experiment.

Given its increasing uptake in the community, experimenters are encouraged to see if LSL is already supported by the devices they use, and investigate whether it would be helpful in their set-ups. LSL provides a compatibility and interchangeability that can make experimental set-ups almost universally plug-and-play.

3.1.3. Research studies in clinical settings

3.1.3.1. MEG acquisition in clinical settings

In the clinical setting, there is often only a single chance to get the acquisition done correctly: a clinician has ordered the procedure, insurance and administration have cleared the patient for their appointment, and the patient has often travelled some distance to arrive for their

exam. If the collection protocol is not designed and followed properly, not only is rescheduling the patient difficult, but the very act of rescheduling places undue stress on the patient, who may assume that some serious finding in their data has necessitated a second exam. Additionally, for pre-surgical studies on patients, there might not be a second chance to correct an erroneous recording — post-surgically, the brain will be completely different. For smooth clinical operations, respect for the patient's comfort, and considerations of costs, we must follow basic minimum standards to ensure a quality exam the first time.

In recent years, several scientific publications have provided guidelines for the collection of MEG data (Bagic et al., 2011; Gross et al., 2013; Hari et al., 2018). Unlike highly-motivated research volunteers, however, patients present their own confounds and contaminants that must be accounted for. More recently, (Mosher & Funke, 2020) provided practical guidelines for the preparation of MEG patients and instruments for routine clinical operations. Here we review key points of the clinical MEG exam, with reference to these other publications for explicit details.

A surprising number of data collection problems can be remedied in post-processing, typically by applying either spatial or temporal filters, but some issues may remain uncorrectable. An example of a generally uncorrectable flaw is bad location data, in other words, incorrect determination and/or positioning of the sensors with respect to the patient. A polhemus 3D localizing stylus is typically used to identify the patient's fiducial landmarks (nasion, pre-auricular points), the patient's head shape, the location of EEG sensors, if present, and — critically for MEG — the location of head position indicator (HPIs) coils affixed to the patient's scalp. The fiducials establish the patient's coordinate system, the head shape data align the structural MRI and/or CT to the coordinates, and the HPI coils locate the patient in the MEG array, all crucial components of subsequent accurate source localization in post-processing.

EEG sensor locations are usually dictated by convention (e.g., most typically 10-20 or 10-10 configurations) typically on "caps" with pre-arranged locations. Certified EEG technicians can accurately and reproducibly locate these landmarks on the scalp. Caps require the selection of the appropriate size for the patient's head circumference, then careful symmetric placement on the head, with the benefit of typically faster application and/or higher sensor densities. In either case, the sensor locations are then determined with a 3D stylus. Commercial MEG arrays are arranged in rigid helmet configurations, which rely on the HPI coils to determine the location of the patient within this helmet space, and thus the HPI coils must be accurately located in the patient's coordinate system. In all cases, the proper use of

a 3D stylus is an art in itself, and therefore technicians must be trained and cross-checked in its use (Mosher & Funke, 2020).

Another critical location component for MEG is to ensure that the patient is as deep in the helmet as possible and remains so throughout the recording. Although perhaps obvious, we are surprised in clinical practice to review data sets where the patient was allowed to “slump” out of the helmet, such that brain coverage by the MEG sensors is incomplete. Recovery of data from cortical locations outside of the helmet is nearly impossible, and therefore the patient must be monitored throughout the recording to ensure that they remain as deep in the helmet as possible (Mosher & Funke, 2020).

EEG sensors are by definition on the scalp, but a similar consideration of coverage is the placement of additional electrodes in the temporal regions and, if warranted, sphenoidal electrodes, to ensure coverage of the temporal lobes and deeper structures. During the recording, both EEG and MEG sensors should be continuously monitored, to detect electrodes that become dislodged or channels that become too noisy or inoperable, so that whole-head coverage remains as complete as possible.

To reiterate, location is the primary consideration to ensure accurate whole-head coverage of the EEG and MEG array. Other considerations covered in more detail in (Mosher & Funke, 2020) include routine degaussing of all patients (particularly useful if an MRI scan was performed recently), changing all patients into MRI666t-compatible gowns, wall charts and “cheat sheets” for technicians to follow setup procedures, and a consistent preliminary signal processing stream.

3.1.3.2. Research studies with intracranial EEG recordings

Intracranial electrophysiological recordings (iEEG), compared to non-invasive techniques, present several important advantages such as enhanced SNR, focal recordings directly from the neuronal source, minimal signal distortion due to the skull and other tissues (Fahimi Hnazaee et al., 2020), and the possibility to record from deep brain structures (Buzsáki et al., 2012; Parvizi & Kastner, 2018). Three main different electrode types can be used for intracranial recordings: strips or grids which are usually sited subdurally (electrocorticography: ECoG), depth electrodes (stereoelectroencephalography: SEEG), and microwires, with higher spatial and temporal resolution that give a precious and rare access to multi and putative single neuron activity in humans. Because of their invasiveness and associated risks of infection and bleeding, human iEEG recordings are exclusively used for

clinical investigations (i.e., brain surgery, deep brain stimulation, investigation of epileptic seizures). Nevertheless, if the patient consents, limited options to acquire additional data for research purposes (pathophysiological or physiological/cognitive studies) are possible and of high added value for researchers, despite the following limitations. The brain coverage of iEEG electrodes is individualized, focal, and sparse, and thus can complicate group studies. The number of patients that can be included in each research project is often limited and hence requires long duration or multicenter studies to accumulate adequate patient numbers. For physiological/cognitive studies, the patient's pathology and medication regimen can affect the recorded activity relative to a healthy participant. Therefore, only selected research protocols can be executed, typically favouring investigations where iEEG has significant and specific benefit relative to non-invasive techniques in healthy participants and are suitable to the patients' abilities (accommodate any cognitive impairments, limited duration of tasks).

Like for all electrophysiological recordings, good recording conditions and data quality is a prime consideration, whether it is for clinical or research purposes. Some aspects are nevertheless specific to research projects involving intracranial recordings in patients. Because of the clinical context, a close collaboration between clinical staff and researchers is key to establishing good coordination of research and clinical activity and collecting high-quality data, especially when dedicated material is necessary like specific electrode type (i.e., hybrid macro-microelectrodes) or amplifier (i.e., with DC recording, high sampling rate, precise trigger time resolution). Optimally, clinicians themselves are involved in the research protocols, and staff are dedicated to interfacing between clinicians and researchers. For recordings with microelectrodes in epileptic patients, all steps from the implantation to the recording require special attention (see detailed recommendations on iEEG electrode manipulation and technical settings in Lehongre et al, this special issue).

Data analysis of iEEG is very similar in many aspects (evoked potential, time-frequency analysis, connectivity, etc.) to EEG or MEG data analysis, but also has some unique procedures. Regarding artifacts, in addition to classical line noise or head and body movement artifacts, there can be electrophysiological activity related to underlying pathology that is being investigated. Intracranial EEG signals are relatively unaffected by blinks or muscle artefacts for deep electrodes; however, if a scalp reference electrode is used, then these recordings can also be tainted with these artifacts. Additionally, it should be noted that iEEG recordings are not immune from volume conduction (Tenke & Kayser, 2012), although expressing activity in a bipolar configuration can aid in highlighting locally generated activity. A very specific intracranial signal analysis is spike sorting. This is used for microelectrode

recordings when action potentials from multiple neurons are recorded and the activity of single cells is studied. Many toolboxes are now available for unsupervised spike sorting (Buccino et al., 2020). Intracranial recordings may also need dedicated tools for signal visualization and electrode localization. Simultaneous recordings of electrodes at different sampling rates (e.g., macro- versus microelectrodes) need specific visualization software to plot activity from all electrodes with different sampling rates on the same window (Ducorps et al., 2010). Electrode localization can be performed by manually scrutinizing a structural MRI or X-ray CT or using automatic detection software that expresses the electrode locations in native 3D space for that particular patient. In the case of research studies where group analysis is necessary, transformation to a common 3D space (e.g., MNI) is required.

Intracranial recordings can be challenging because of the number of additional investments (technical, financial, temporal, personal) and some limitations. Nevertheless, they undoubtedly provide unique added value for exploring human brain function, giving direct access to neuronal activity and even single cell action potentials, currently unrealizable with non-invasive techniques.

3.1.3.3. Simultaneous intracranial EEG and MEG/EEG

As noted above, iEEG electrodes minimally distort brain signals but nevertheless have limited sampling capability. In contrast, MEG and EEG provide information on overall brain activity from sensors typically covering the scalp with: (1) lower SNR than iEEG (often compensated through averaging (Puce & Hämäläinen, 2017)); (2) effects of volume conduction and spatial smearing from scalp and skull in EEG recordings (Hari & Puce, 2017); and (3) requiring a solution to the “ill-posed” inverse problem (i.e., with a non-unique solution) to localize activity in grey matter (Baillet et al., 2001).

Therefore, simultaneous recordings of iEEG, scalp EEG and/or MEG capitalize on the strengths of each respective technique to describe the underlying brain activity accurately (high SNR) and comprehensively (overall view). In particular, understanding the relationship between surface signals and the spatiotemporal configuration of the underlying cortical sources is central to both basic fundamental and clinical research. First, in basic research, iEEG has sometimes been considered as capturing the “ground truth” for evaluating how known cortical activity is captured by non-invasive methods. For instance, some studies using simultaneous intracranial and scalp recordings have challenged the assumption that activity arising from deep sub-cortical sources such as the hippocampus does not exhibit a sufficient SNR to be detected with MEEG surface sensors (Koessler et al., 2015; Pizzo et al., 2019; Seeber et al., 2019) without the need for excessive averaging (Fahimi Hnazaee et al.,

2020). Second, in clinical settings, simultaneous iEEG, scalp EEG, and/or MEG recordings can provide complementary information regarding the epileptogenic zone (Gavaret et al., 2016; Kakisaka et al., 2012; Santiuste et al., 2008). For example, Gavaret et al. (2016) used epileptic spikes detected in iEEG, but not visible in scalp EEG, to reconstruct the overall brain activity using MEG signals. In this study, the multimodal approach clearly showed a posterior generator that was not covered by the iEEG implants. Finally, in systems and cognitive neuroscience research, the analysis of signal fluctuations evoked by individual stimuli in depth and surface recordings can be exploited as a crucial source of information for refining our understanding of the neural activity underlying specific cognitive processes (Dalal et al., 2009; Dubarry et al., 2014).

However, introducing scalp EEG, and/or MEG simultaneous recordings to an already highly-constrained recording setup for iEEG creates a large number of important considerations related to quality data acquisition and GSP.

For scalp EEG, the placement of external electrodes can be challenging because of the presence of the surgical dressings. While this setup allows for recording at the patient's bedside for a prolonged period of time, it requires a high level of maintenance with constant regelling of scalp electrodes and addition of physiological adhesive to ensure that scalp electrodes remain in place and operational. In such a setup, signal analysis must also take into account the dramatic changes in electric field propagation caused by the implants, their burrholes and craniotomies, which create large skull discontinuities (Dalal et al., 2009; Kirchberger et al., 1998). For MEG, conducting simultaneous recordings with iEEG requires dedicated equipment and space (i.e., an MEG facility housed in a clinical setting). The complex setup presents obvious physical constraints (e.g., fitting bulky iEEG electrode connectors within the MEG dewar) and the exam must be carried out within a limited period of time (e.g., usually not more than a one-hour exam). It is worth noting that in such setups MEG signal quality can be drastically affected by the presence of the iEEG electrodes and equipment (e.g., cables, connectors). To date there is only one study that has achieved the simultaneous acquisition of MEG, EEG and iEEG signals (Dubarry et al., 2014), because of the significant technical challenges that occur when introducing metallic materials into a shielded MEG environment. Additionally, such recording sessions require complex logistics and organization between clinical staff, patient, technologists/research engineers, and clinicians to ensure smoothness of procedures and minimization of time spent outside of the epilepsy monitoring unit, as there is a risk that seizures might not be recorded optimally on simultaneous video and EEG.

In summary, the simultaneous recording of surface and depth neurophysiological signals provides a unique perspective on various aspects of brain activity, with potentially significant benefits for clinical and research purposes. However, data collection is technically very challenging, from obtaining patient consent (explaining the procedures) to data acquisition, curation, analysis, and interpretation. Therefore, planning such recordings should only be considered when they are clearly motivated by the research hypothesis and strongly supported by the clinical staff.

3.1.4. Open MEEG data repositories

Data sharing is essential for solid scientific progress (Gorgolewski & Poldrack, 2016; Poldrack et al., 2017, 2020). While significant resources have been invested in neuroimaging studies worldwide, the number of projects which publicly share their data remains limited. Over the last years, some initiatives have emerged to openly share MEEG datasets. These repositories foster reproducibility of results and help to answer specific explorative scientific questions not originally conceived at the time of data collection and study design, increasing data usage and longevity. Furthermore, they allow data aggregation attaining larger sample sizes which could be used to increase statistical power and approaches involving artificial intelligence. Open repositories also facilitate data access to researchers with less funding opportunities and resources, giving back to society part of the value invested. Some examples of open MEEG repositories are: Human Connectome Project (HCP) (Larson-Prior et al., 2013), The Open MEG Archive (OMEGA) (Niso et al., 2016), The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) (Shafto et al., 2014; Taylor et al., 2015), The Temple University Hospital EEG Data Corpus (TUH EEG Corpus) (Obeid & Picone, 2016), OpenNeuro (Markiewicz et al., 2021). See more resources in Table 1: Resources for GSP.

3.1.5. Simulating data

Up to this point we have discussed data obtained by recording brain activity from actual (human) participants/patients. However, there are a number of reasons why the generation of so-called synthetic, toy, or simulated data may sometimes be preferable to data from live participants.

One simple, practical reason for using simulated data concerns the cost of data collection. Recording real data requires development and validation of specific experimental paradigms, time to record from a sufficient number of participants, and the acquisition of all necessary supportive material including human resources. For many researchers, planning and

conducting experiments are the most cost-intensive activities they perform. Whenever simulated data can fulfill the same purposes as real data, e.g., when actual human cognitive processes are not the main focus of the study, simulating data can be much more efficient than recording real data, especially given the dedicated simulation tools available today. This can be particularly useful when developing new analysis methods, or when testing planned analysis pipelines prior to implementing actual data collection.

Another reason why it may be preferable or even necessary to use simulated data instead of real data is that only simulated data can provide a fully known ground truth. This ground truth is needed to verify the accuracy of any method, i.e., the extent to which the results reproduce this known truth. Unfortunately, with MEEG recordings no such ground-truth ideal of brain activity is available. Therefore, researchers often construct synthetic data in such a way that its ground truth is known, allowing the idiosyncrasies of the method to be properly evaluated.

A number of software packages have recently become available for creating simulated data via a number of fundamentally different methods. For example, generative-adversarial neural networks can mimic existing EEG data (Hartmann et al., 2018). Such neural networks take previously recorded real data and generate new synthetic data that cannot be distinguished from the original data it is based on. As such, the resulting data does not contain a known ground truth, but can be used for data augmentation, data recovery, or up-sampling. The Virtual Brain, on the other hand, uses interconnected neural mass models based on differential equations to simulate the low-level dynamics of clusters of neurons (Sanz Leon et al., 2013). This results in a potentially highly detailed ground truth. However, the most common method of EEG data simulation employed in the past decades relies on a simplified forward model which describes EEG as a combination at the scalp of activity projected from a limited number of generally independent sources in the brain. This is captured by the equation $x = As + \epsilon$, where x represents the scalp EEG data, s the underlying source activations, A a matrix projecting the source activations to the electrodes, and ϵ an error term which captures any remaining noise. Following this equation, data simulation thus requires the generation of a number of source activations, the linear mixing of these source activations to simulate scalp activations, and the optional addition of noise. By generating these three factors themselves, scientists can be sure to know the ground truth of the data. Previously, researchers often wrote their own code for such simulations, limiting standardization and reproducibility. However, one recent free and open-source MATLAB-based EEG data simulation toolbox, SEREEGA, is based on this common method

and provides a standardized, reproducible framework to simulate data in this fashion (Krol et al., 2018). See Table 1: Resources for GSP for a selection of other MEG simulation tools.

Note that any simulation approach is, itself, a method that makes certain assumptions. As such, it is a tool that provides certain functionality, but cannot guarantee any particular quality or appropriateness of the outcome. Each simulation needs to be performed for a particular purpose, requires different parameters, and makes specific demands with respect to different kinds of validity. As is the case with all tools and methods, it remains crucial to carefully consider a simulation's parameters and their appropriateness for the particular task at hand. For example, when a method that assumes linearity in the data performs accurately on simulated data that has been constructed using a linear model, it has not been proven that this method will transfer well to real data. That said, simulation is a powerful tool, as it gives us full control over the “data”. An appropriate simulation can provide a standardized way to objectively and reproducibly validate and evaluate the tools we develop and use.

- → Careful attention to data collection is crucial as conditions during data collection can affect data quality, with different contexts and different modalities having different requirements.
- → The Lab Streaming Layer can help reduce complexity in multi-modal set-ups.
- → In clinical settings, the unique constraints require special attention to, i.e., localization and the cooperation between several actors.
- → Combined modalities help gain new insights, but come at the cost of more complex setups, and additional constraints.
- → Open repositories promote results reproducibility, data re-usage, increasing long-term value
- → Sometimes a project is best served by using simulated data instead of real recordings, e.g., when a known ground truth is needed.

3.2. Analysis software strategies for GSP

Computer software is at the core of any MEEG data analysis project. Be it code that scientists write from scratch, or software toolboxes that support them, GSP is highly valuable at this level too, if reliability is to be promoted. In this section, we first highlight a few key principles that scientists can benefit from when designing code, and then briefly review how today's most widely used software toolboxes have developed and engage in GSP.

3.2.1. Designing code for correctness and reproducibility

A core ingredient of a scientific experiment is to document, in detail, what steps have been performed, so that they may be reproduced. When the number of steps is small, the Methods section of an article may contain them in full. However, journal page limits ensure that this is frequently not the case in modern MEEG studies, requiring Methods to be summaries that omit many important details so as to not cloud the 'big picture'. This leaves the programming code itself as the authoritative source of the exact analysis procedures used in the collected data. With modern programming languages and tools, code can serve this purpose well, if *it is consciously written to do so*. If, however, the programmer's focus is solely on obtaining the desired output, without regard for transparency in how results are obtained (e.g., documentation, easy to follow program flow, version control), chances are high that the results will not be reproducible, or maybe even wrong (Casadevall et al., 2014; Miller, 2006; Pavlov et al., 2021).

Much has been written already on the topic of how to produce code that clearly communicates analysis steps, and is applicable to stimulus presentation and experimental control scripts. The reader is encouraged to look at the writings of (van Vliet, 2019) and (Wilson et al., 2014). The most important thing is to have the mindset that scientific code *is not done until it is well organized*. This has little to do with technical skills. Rather, this is about proper "housekeeping": making sure unneeded code and files are archived and removed from the main project, making sure everything is named properly, long and convoluted scripts are cut up into commented modules that are easier for others to understand, etc. As the analysis advances and new code is written, clutter inevitably accumulates. Fortunately, there are tools for organizing files containing code, such as version control systems (see Table 1: Resources for GSP). With some discipline, these tools allow keeping track of every change made to the code, and allows efficient collaboration at scale. Furthermore, more and more analysis toolboxes such as those presented below offer script generators to improve reproducibility and generate code that is easy to understand and adapt (Es et al., 2021), which reduces the burden on the researcher.

3.2.2. Software toolboxes for stimulus presentation

There is a wide array of software solutions for stimulus presentation and response registration, including online platforms. Since even the same software may perform differently on different hardware with respect to latency and jitter, the main GSP recommendation here is to properly measure and document these parameters. This can be achieved via calibration scripts that display simple stimuli at known timing and spatial

locations. Precise onsets and offsets of these stimuli can be measured by connecting an oscilloscope to: (i) the computer port presenting the stimuli; and (ii) the physically presented stimuli themselves, either via photodiodes placed on the computer monitor, or microphones placed in front of speakers, providing a measure of the precision of the stimulus presentation software. Once the delay and jitter are known and documented, it may be necessary to adjust the analysis pipeline to take this into account (see also section 3.1.2. Multimodal data collection above). Calibration test results represent research data in their own right, and the results of teams using different combinations of hardware and software can provide valuable information about optimal configurations for certain kinds of tasks. This is particularly important in experimental protocols being implemented in multiple testing sites. To make optimal use of this information, software and hardware configurations need to also be documented in detail, as discussed in section 3.6 Current efforts for data standardization, analysis, and reporting.

3.2.3. Software toolboxes for data analysis in MEEG studies

A number of comprehensive software toolboxes exist for analyzing MEEG data. The choice of an analysis environment to work in should not be made lightly, as this decision may commit the data analyst for not only the current project but also future studies, should reproducibility be a specific priority. Some points to consider are the required (and available) level of programming skill and financial resources, the availability of a graphical user interface, and the availability of local software support and expertise. However, ultimately, not all toolboxes are equally well-suited for all types of analyses or modalities (EEG, MEG, fMRI…). Although the most common analysis methods are implemented in most toolboxes, some offer more advanced algorithms in one or another environment, and more advanced projects may require applying methods available across several toolboxes. In that case, manipulating data across environments and formats will be necessary, and extreme care should be taken in assuring that all technical information is handled properly across environments. Monitoring and seeking advice from the community via the relevant user forums is advisable.

Table 2: MEEG analysis Toolboxes highlights current differences and summarizes the most relevant features of five of the most commonly used MEEG analysis environments to date. These toolboxes are under constant and active development, and the developers cooperate to reimplement each other's methods and reflect on the best ways to do so (e.g., Jaiswal et al., 2020, or *From Raw MEG/EEG to Publication: How to Perform MEG/EEG Group Analysis with Free Academic Software.* | *Frontiers Research Topic*). These toolboxes all implement

standard and advanced analyses, and enforce GSP in some way, but each with their own peculiarities, inherited from the laboratory traditions, equipment, software environments and applications for which they were developed (see Box 2). Attention to these specifics may aid the decision for which toolbox(es) to optimally use for a given project.

These toolboxes can generate scripts and reports – in line with current GSP. In particular, reports that mix code and human readable narratives are handy tools that allow saving the results and the code that generated them in one document, easily accessible, and readable with any document viewer. If these reports expose all (including default) parameters and software versions, they can readily be used to reimplement an analysis with minimal burden, even in a relatively distant future. This feature is particularly important because backward compatibility with previous versions can hardly be ensured. At the time of publication, archiving of code and reports on to a public repository, ideally with a permanent digital object identifier (DOI), is a desirable step that greatly facilitates reproducibility.

A final, general point of GSP concerning toolboxes is the following. Even if a single toolbox is sufficient for a given project, it can be worthwhile to perform the analysis across multiple toolboxes. Each toolbox is likely to have its own idiosyncrasies, its own unique implementation even for ‘standard’ methods. To avoid results being affected by such issues, it is GSP to perform the same analysis pipeline in different toolboxes and compare their outcomes. When the same analyses across toolboxes support the same conclusion, this strengthens the robustness of the results.

- → Code is often the only authoritative source of exact experimental procedures, so well-written code greatly aids reproducibility as well as correctness.
- → It is important to verify the visual and temporal accuracy of the stimulus presentation, since this is often the main component in an experiment.
- → Choosing a software environment to analyse MEEG data requires considering the local lab environment, programming capabilities, and the needs of the experiment being performed.
- → Toolbox environments evolve over time, and different implementations of the same procedure can have different outcomes. It is important to document the precise version of any software package or coding environment, and to verify results in different toolboxes when in doubt.

3.3. Signal processing

Just like all the other steps along the research lifecycle, from experimental design to result interpretation, the signal processing step can make or break a study. Signal processing is the step that extracts from the raw data the key information needed to answer the scientific question being asked. The nature of the pitfalls or limitations, and the recommendations to avoid them, vary largely depending on the specifics of the methods and characteristics of the recordings. Therefore, given the ever-growing number of tools routinely used to (pre)process MEEG data and the wide range of experimental protocols in the field, a systematic exploration and exhaustive account of GSP for all MEEG signal processing is not practically feasible (although a number of guideline documents exist, listed in the introduction). Instead, this section offers general high-level recommendations for better planning and designing the signal (pre)processing components of a MEEG study.

First, the goal of the study (i.e., the question it seeks to resolve) dictates a possible set of signal processing methods. From this set, the analysis methods should be chosen *before* finalizing the experimental task design (e.g., via 2.2. Pre-registration). This ensures that collected data can be optimally examined with the projected tools. For example, if the plan is to examine slow oscillatory components in the signal, or run detrended fluctuation analysis on the data, it is crucial to make sure the temporal windows that will be analyzed are long enough to allow for such analyses, or that the data themselves are collected using appropriate analog filtering. Although it is possible to consider one's analysis options based on the properties of the available data, this approach might result in "p-hacking" and other biases mentioned in previous sections. It is therefore strongly recommended to proceed *the other way round*: First decide which method to use to address a research question, and then create an adapted experimental design to comply with the methodological constraints.

Second, it is very important to understand how different signal processing tools transform the data, and what the transformed signal reflects or measures. This does not mean that everyone should master the mathematical and implementation details of the signal processing tools they use, but it is important to develop an *intuitive understanding* of these tools and to be aware of the parameters and default settings associated with the various processing steps in the pipeline (e.g., what filter parameters were used and why). The minimal level of understanding of a tool should include an appreciation of its limitations or pitfalls. A case in point is the importance of understanding the limitations and specificities of distinct connectivity metrics, e.g., problems associated with metrics sensitive to zero-phase

lag coupling, and even potential limitations that arise with methods thought to overcome these very issues (Palva et al., 2018).

Moreover, an issue that is sometimes overlooked is the feasibility of connecting various processing steps into one's signal processing pipeline, and the order in which multiple operations are performed. Simple examples include the importance of computing time-frequency maps in single-trial data before averaging across trials when investigating induced responses (Tallon-Baudry & Bertrand, 1999). Likewise, it is important to verify whether certain pre-processing steps (down sampling, filtering, ICA etc.) preclude the subsequent application of specific signal processing tools (Cohen, 2014).

In addition, when choosing certain methods, it is important to consider implicit assumptions about data properties and whether these are warranted. As an example, choosing to only compute power in a given frequency band (e.g., theta-band) without examining the full power spectrum is usually based on the assumption that the data actually exhibit specific oscillations at that frequency. This assumption can be confirmed by examining the entire spectrum for a peak at the frequency range of interest. As a matter of fact, several methods can verify the presence of oscillations and distinguish them from the background 1/f component of MEEG data (e.g., EMD (Quinn et al., 2021), FOOOF (Donoghue et al., 2020), or Samaha et al, this issue).

Furthermore, similarly to replicating results using different toolboxes (as mentioned in section 3.2.3. Software toolboxes for data analysis in MEEG studies), conducting the same analysis using alternative signal processing methods can be a useful approach to test for robustness of an observed effect and potentially to troubleshoot analyses. Obviously, attempting to replicate (or compare) findings using multiple methods requires time and effort, and it cannot be done for all analysis steps. However, it can be a useful practice that can be applied to a few key steps in the pipeline. Examples include using more than one source estimation technique, or using both wavelet-based and Hilbert-based implementations of spectral analyses metrics. Likewise, exploring a range of values for the parameters of a given analysis rather than sticking to default values can be a way to probe the robustness of the results as a function of parameter settings. If small changes in the parameters lead to substantial changes in the results, one should not systematially dismiss the results as unreliable, but rather view this as a warning call for further investigation.

Another important habit to develop in connection with signal processing is frequent data visualization. As often as possible, the output of intermediate analysis steps throughout the

signal processing pipeline should be plotted and carefully examined. This can be achieved by using modular coding approaches with opportunities to examine the output of each signal processing stage (see also section 3.2.1. Designing code for correctness and reproducibility). This also facilitates rapid trouble-shooting and in some cases might even trigger new ideas for data analyses.

The recent proliferation of studies using machine learning (ML; e.g., brain decoding) has generated a lot of interest and excitement in MEEG. However, the rise of "data-driven neuroscience" should not occlude the fact that assumptions made about the data and the implicit knowledge of the experimental paradigm both dictate analysis choices and/or parameter tuning (e.g., choice of extracted features, task design, or target classes in decoding studies using supervised learning, etc.). Data-driven approaches are not more "objective" methods, and their rise is not a sign that hypothesis-driven neuroscience has come to an end: both approaches complement each other. The best practices and pitfalls of MEEG signal processing in the context of ML analytics are largely the same as those that apply when exploring data from other domains. Among other things, one should watch out for violations of the strict separation between training and test data (e.g., avoid analyses that induce dependencies across the whole dataset). Section 3.5. Artificial inteligence for MEEG data analysis presents more recommendations for the use of artificial intelligence (more precisely, machine learning) for MEEG data analysis.

- → The method used to answer an experimental question sets specific data requirements and should be chosen before designing the experiment.
- → Data analysts should understand all signal transformation steps, assumptions, limits and pitfalls of the used methods.
- → Running the same analysis with several signal processing alternatives and visualizing the intermediate analysis steps can help assess the robustness of experimental results.
- → Scripting and documenting analyses is essential for reproducibility.

3.4. Statistics

In experimental MEEG data analysis, statistics are mostly considered as an inference tool. Given a particular experimental design, statistics help experimenters decide whether or not a given treatment affects measurements, or if two groups have sensibly different signals in a given region of interest (electrode, time window, frequency band, anatomical location…). It is one of the challenges of GSP to keep these inferential statistics valid across the various

situations in which they are used. Other, less obvious uses of statistics are nevertheless quite common in MEEG data analysis. We now report on three statistical aspects that are important for GSP.

3.4.1. From hypothesis-driven to data-driven approaches

Over the last few decades, statistics have become a tool for data-mining in exploratory (or so-called data-driven) analysis and whole-brain inference, where mass univariate statistics (Groppe et al., 2011) and multivariate analysis (Cichy et al., 2014; King & Dehaene, 2014) deal with the challenges of multidimensional MEEG data. These new use cases have stimulated a number of changes in GSP. A well-known example is the adaptation of decision thresholds to the number of tests being performed (the so-called multiple comparisons problem). Performing tests across the entire data space increases the chances of reporting false positives. Methods such as family-wise error rate (FWER) or false discovery rate (FDR) prevent this type of inflation. Among those methods, non-parametric permutation-based approaches (Maris & Oostenveld, 2007; Smith & Nichols, 2009; Sassenhagen & Draschkow, 2019) currently outperform other methods (Hayasaka & Nichols, 2003; T. Nichols & Hayasaka, 2003; Puoliväli et al., 2020). For decoding approaches and out-of-sample generalizability, permutations have also been used for a data-driven robust estimation of "chance level" (Combrisson & Jerbi, 2015), adapting for choices of cross-validation schemes (i.e., how to split testing and training sets) and decoders. For a comprehensive introduction to nonparametric (permutation-based) statistical testing in MEEG, see (Maris & Oostenveld, 2007).

3.4.2. Simulations for prospective power analysis

It was recently argued that neuroscience studies are often underpowered (Button et al., 2013; Ioannidis, 2005; Szucs & Ioannidis, 2017), resulting in inflated and unreplicable effect sizes. To avoid this, actual effect sizes are currently being measured, and registered with projects such as ERP CORE, or #EEGmanylabs (see section 2.3.3. #EEGManyLabs), and their detectability evaluated. For the latter, simulations are important. As already noted (section 3.1. Data collection), simulations allow evaluating the chances of detecting a given effect with statistical tests, and hence to assess statistical power. Ensuring that proper power is achievable in a given setting is essential to guard against wasting resources on experiments with no chance of demonstrating effects, even if they exist. In a recent simulation-based study, (Chaumon et al. 2021, this issue) examined how the expected spatial properties of sources of MEG activity affect statistical power. This study clearly

showed that spatial variability in the source of the signal and the type of contrast measurements strongly affect statistical power. Taking into account individual anatomical variability of expected active regions in a given dataset is key in experimental MEEG studies, since the number of samples (trials and participants) that are required to achieve a given level of statistical power vary several fold between areas with lower spatial variability (e.g., in the precentral sulcus) compared to regions with higher spatial variability (e.g., lateral occipital cortex). Specific MEEG simulation toolboxes (Krol et al., 2018), or dedicated statistics simulation toolboxes (Lakens & Caldwell, 2021) can be used for prospective power analysis. If hypothesis testing is conducted under a Bayesian approach, simulations can help set a maximum sample size for a Sequential Bayesian Factor design. In this design, the researchers run their analysis pipeline after data from each participant (or batch of several participants) are collected. Data collection ceases when there is strong evidence for either H1 or H0, or when the maximum sample size has been reached (Schönbrodt & Wagenmakers, 2018).

3.4.3. Quantifying data quality

A fundamental challenge in MEEG research – where signals are typically tiny relative to noise – is to obtain precise, reliable measurements of the brain signal of interest. Remarkably, however, the field does not have a standard and widely-used approach to quantify data quality in a given study. In most published papers, there is simply no way to objectively evaluate how “clean” the data are. Two approaches have been developed recently to address this issue in the context of averaged event-related electrical activity (but could easily be applied to magnetic data).

One of these approaches takes traditional metrics of “reliability” from psychometrics and applies these metrics to ERP amplitude or latency measures (e.g., (Olvet & Hajcak, 2009; Pontifex et al., 2010)). In psychometrics, reliability is typically defined as the proportion of the total variance across participants that is the “true score variance” (true differences between participants, as opposed to differences due to noise in the data). This approach is particularly useful in individual differences research, which asks how well a neural measure correlates with some other measure (e.g., with a symptom score). The correlation between two variables is straightforwardly limited by the psychometric reliability of the individual variables. Thus, a MEEG measure must be reasonably reliable to be used in correlational analyses. This approach has two limitations, however. First, it provides a single reliability value for an entire group of participants and provides no information about the data quality for individual participants. Second, the reliability value depends on the amount of true score

variance present in the group of participants, making it difficult to generalize across subject populations.

The second metric for quantifying ERP data quality is called the *standardized measurement error* (SME), and it is an extension of the general concept of the standard error of measurement (Luck et al., 2021). The SME value for a given participant quantifies the precision of an ERP amplitude or latency score for that participant (i.e., the extent to which you would expect to obtain a similar score if you repeated the experiment multiple times for that participant). Unlike psychometric reliability values, the SME is computed individually for each participant. However, SME values can be aggregated across participants to quantify the overall data quality of a given experiment and to estimate the impact of noise on the effect size and statistical power of the experiment. The SME can also be combined with a new power calculator that takes into account the number of trials as well as the number of participants (Baker et al., 2020), making it possible to predict the effects of changing the number of trials in future experiments with the same basic paradigm. A basic version of the SME is now automatically provided when averaged ERPs are created by ERPLAB Toolbox (an EEGLAB plugin; Lopez-Calderon & Luck, 2014), and example code is provided for more complex applications (Stewart & Luck, 2020). An important limitation of the SME is that it is limited to amplitude and latency measures obtained from averaged waveforms and cannot be applied to single-trial analyses.

- → Reference to statistics is often restricted to inference, but this wide field of research offers tools for many more applications
- → Current uses of statistics for univariate and multivariate data-driven approaches require adapting significance thresholds
- → Simulations can help prospective statistical power analyses
- → Data quality can be evaluated in acquired data

3.5. Artificial Intelligence for MEEG data analysis

Since its introduction by Alan Turing (Turing, 1950) and formalization in the 1950s, the transdisciplinary field of Artificial Intelligence (AI) has kept growing to impinge on many disciplines, including neuroscience. The transdisciplinarity of AI combined with the complexity of physiological signals inevitably leads to a broad range of misunderstandings, misinterpretations, uses, and misuses in different fields (Wang, 2019). Here, we focus on AI

as a tool for processing and analysing MEEG data, notably through one of its subfields: Machine Learning (ML). ML refers to applications of AI in which algorithms build models to make predictions or decisions "[...] without being explicitly programmed" (Koza et al., 1996). A survey of 1.3 million openly available research articles revealed that, while classical statistical methods are still widely used, multivariate statistical and ML approaches have seen a significant increase over the last decade (Bolt et al., 2021). See, e.g., (Lemm et al., 2011) for an introduction to ML for brain imaging and Scikit-learn (Pedregosa et al., 2011) as a widely used resource to perform ML analysis.

Algorithms used in data-driven AI methods, such as ML, use all data they are provided with to optimize their outcome – regardless of their (neurophysiological) nature or any commonsense knowledge. These algorithms cannot understand the data as humans can (Rusu et al., 2016). Therefore, if the data contain noise, an ML algorithm may use it in unpredictable ways, leading to unreliable or even false outcomes (Hedderich & Eickhoff, 2021; Roberts et al., 2021). As a result, data quality becomes extremely crucial, and experimental designs with adequate signal-to-noise ratios are needed. The quality of the data output from the algorithm cannot be better than the quality of information coming in or, as the old saying goes, "Garbage in, Garbage out" (Babbage, 1957).

An important point about noise in biosignals like MEEG is that it is not random: it may also be systematically correlated to the conditions of interest. The AI algorithm uses this noise to improve its performance outcome and creates seemingly reliable results, which can be nonsensical for the neurophysiological effect one is trying to characterize. This has been the case in the Brain-Computer Interface community (Roy et al., 2019), where several algorithms have reached very high classification rates thanks not to brain activity, but due to muscle and/or ocular artifacts present in the same data. These artifacts are deemed more relevant by the classifier than the actual brain activity. It demonstrates that the provided dataset was biased by more motion and artifacts in one condition compared to the other(s), and that (inappropriate) data pre-processing had not removed these artifacts.

The above examples stress why signal (pre-)processing pipelines should also be used with a good understanding of their basic assumptions and effects on data so as to keep ML-based algorithms prerequisites fulfilled. If we take the very basic example of dataset slicing (mentioned in section 3.3.), some conditions have to be met (e.g., the test set size ensuring statistical relevance, or the independence kept between the training and testing sets). Yet, some paradigms (e.g., block designs (Lemm et al., 2011)) or data processing techniques

introduce dependency between samples or across the whole dataset, thus violating these assumptions.

AI-based methods can also produce unstable outcomes when they do not have enough data. Notably, deep learning-based EEG analyses (Roy et al., 2019) require a substantial amount of data, but often available datasets usually contain too few tasks, participants, sessions, samples, etc. to ensure powerful and statistically relevant analyses (Hinss et al., 2021). Furthermore, including a larger sample and more diverse participants – as well as datasets from a broader set of recording environments – allows preventing the so-called "algorithmic injustice" (Birhane, 2021), whereby the analysis may lose in generalizability and interpretability in smaller, likely more biased, samples. This again highlights the importance of data sharing for more meaningful results with MEEG data analysis.

Last but not least, there is the issue of interpretability. The increasing ease of use and availability of ML algorithms makes it increasingly likely that their results may not be interpreted correctly. This difficulty of interpretation is aggravated by a lack of transparency of algorithms which can be black boxes and prevent from apprehending the nature of their outcomes. There is no point in attempting to interpret outputs of mathematical functions which have no neurophysiological meaning (Haufe et al., 2014). To evaluate the actual neurophysiological basis of the results and make sure that the results are reliable, such interpretable methods are invaluable.

One project that deals with the above-mentioned issues is the Mother Of All BCI Benchmarks (MOABB; see Table 1: Resources for GSP), which intends to promote valid and reproducible BCI research through algorithm availability, benchmarking, ranking, and freely available datasets.

- → AI is a transdisciplinary field with increasing applications for neuroscience, in particular for data analysis.
- → AI algorithms will always provide an outcome, but its relevance and accuracy depends on the quality and meaning of the input data, an experimental design taking into account known AI pitfalls, accuracy of input parameters, and neurophysiological interpretations of the produced models.
- → Sharing data and algorithms will improve generalizability and representativeness of AI algorithms as well as their accuracy, and prevent demographic biases.

3.6. Current efforts for data standardization, analysis, and reporting

As mentioned in the introduction, one aim of this paper is to foster a general understanding of GSP that goes beyond previously established best practices in this era of Open Science. Here, we discuss a small selection of current standards that provide a more specific, hands-on approach specifically for data organization and methods reporting.

Currently, storage and organization of data is largely performed in an idiosyncratic manner, sometimes differently even within the same lab. When all data are organized in a consistent and structured way, different analysis methods can have standardized access to it, sharing and documenting it becomes easier and less prone to errors. Importantly, it benefits the experimenters themselves when they revisit their data at later points in time.

Across laboratories there are countless potential recording, pre-processing, and analysis pipelines for a given type of study, leading to the “garden of forking paths” (Gelman & Loken, 2013), as has recently been documented in a case analysis of the N400 ERP component (Šoškić et al., 2021) (see section 3.6.2. Reporting results: ARTEM-IS). Of 132 analysed N400 papers, each reported an idiosyncratic approach to acquisition, pre-processing, and analysis, with some methodological decisions being so diverse that almost no two studies took the same approach. What is particularly problematic is that it is typically not possible to fully reconstruct which pipeline has been implemented in a given study, as journal articles generally do not include complete and detailed MEEG methods descriptions (Clayson et al., 2019; Šoškić et al., 2021).

The inability to fully reconstruct all details of an MEEG study would not be a pressing issue if results were robust to variations in the recording, pre-processing, and analysis pipeline. However, a number of studies that have varied methodological decisions to compare outcomes have all concluded that this is not the case. For example, it has been shown that sample size (Boudewyn et al., 2018), EEG recording systems (Melnik et al., 2017), electrode impedance (Kappenman & Luck, 2010), filters (Tanner et al., 2015), statistical analysis (Luck & Gaspelin, 2017), and other methodological decisions (Sandre et al., 2020; Šoškić et al., 2019) can all affect study outcomes.

Although many published guidelines on good scientific practice (see section 1. Introduction) include descriptions on accurate reporting, these guidelines are not sufficient to solve the reporting challenge. It is for these reasons that several initiatives have emerged in the neuroimaging community to deal with dataset organization, description, and reporting, three of which are introduced in this section. Unlike earlier attempts at standardizing reporting on

MEEG studies, these new initiatives have taken the form of living documents to keep up with ever-evolving perspectives of GSP, as well as the development of new tools and analyses. Additionally, these tools all make use of the more recent option to provide rich supplementary materials to help researchers achieve the level of clarity that cannot easily be met through methods descriptions in journal articles.

Prior to manuscript publication in a scientific journal with formal peer review, sharing a preprint version of the document on a public open server such as bioRxiv, medRxiv, arXiv, NeuroLibre, etc., allow broader and faster accessibility of scientific results, skipping paywalls and fostering free dissemination worldwide.

3.6.1. Data organization: BIDS

The Brain Imaging Data Structure (BIDS) is a community-led standard for organizing, describing, and sharing neuroimaging data (Gorgolewski et al., 2016). This standard facilitates data sharing and the development of analysis tools in the neuroimaging community. As an evolving standard, BIDS already supports multiple neuroimaging modalities including MRI (Gorgolewski et al., 2016), MEG (Niso et al., 2018), EEG (C. R. Pernet et al., 2019), iEEG (Holdgraf et al., 2019), and PET (Norgaard et al., 2021). The BIDS specification includes specific details for folder and file naming, the choice of data formats, and the representation of metadata. Multiple accompanying data examples, applications, and tools make it easy for researchers to incorporate BIDS into their current workflows, maximising reproducibility, enabling effective data sharing, and ultimately supporting good data management practices (see Table 1: Resources for GSP). In particular, there are numerous tools to deal with MEEG data: 1) *BIDS converters:* different tools are included in commonly used MEEG analysis packages to convert MEEG data and metadata into BIDS; 2) *general tools for data querying and related operations*, e.g., PyBIDS (Yarkoni et al., 2019) and BIDS-Matlab; and 3) *BIDS analysis tools*, e.g., BIDS Apps, which are containerized analysis pipelines that take a BIDS-formatted dataset as their input and produce derivative data (Gorgolewski et al., 2017). BIDS is an open and inclusive community, and new members may start contributing to the initiative through the BIDS Starter Kit (see Table 1: Resources for GSP for links to useful resources).

Presently, the BIDS Common Derivatives specification is under active development, setting the principles for organizing and describing outputs of processing pipelines. Thus, broadening BIDS standard beyond “raw” neuroimaging data. The specification of BIDS derivatives (i.e., outputs of processing pipelines) capture data and meta-data sufficient for

researchers to understand and critically reuse those outputs in subsequent processing. BIDS has the advantage of being both human and machine-readable, facilitates data interaction by researchers, and accelerates tools development.

3.6.2. Reporting results: ARTEM-IS

As described above, achieving the level of detail sufficient to adequately replicate an MEEG study has proven to be a challenging task, despite the availability of guidelines and checklists designed to facilitate it. For this reason, a more streamlined approach has recently been suggested in the form of metadata templates which require precise numerical/categorical data to be filled. Such metadata templates not only ensure a level of clarity in reporting that is difficult to achieve in a verbal description, but also provide a standardized data format for searchability and simplicity in future metascience, data sharing and reuse (Gau et al., 2019; Styles et al., 2021).

In addition to developing guidelines and user-friendly checklists, systematic reviews of the existing literature are crucial for identifying existing reporting issues and errors in the literature. The same study cited earlier analysing the methodology of 132 N400 papers (Šoškić et al., 2021), found that 100% of papers contained at least some omissions from their methodological report, and 46% of papers contained at least some ambiguities in their described procedures. Particularly poor reporting was observed for properties of filters applied both during and after recording of EEG, and data reduction (i.e., the number of trials remaining for analysis after behavioural and signal-related exclusions). Furthermore, the order of operations was unreconstructable in around half of all cases, and ambiguous in the majority of the remainder.

The data items identified in this review were used to create an evidence-based template that can be used either to document what methods details have been reported in published research or to document the details of current research in progress. The resulting template is a living, version-controlled document called ARTEM-IS: an Agreed Reporting Template for EEG Methodology - International Standard (Šoškić et al., 2020; Styles et al., 2021). By its design, the ARTEM-IS template requires data to be entered in a particular format, thereby improving the accuracy of documentation and standardizing the format of the reported details. At the time of LiveMEEG (October 5-9, 2020), the ARTEM-IS template consisted of 93 fields designed to capture ERP methodology up to the point of statistical hypothesis testing, in a loosely formatted spreadsheet (Styles et al., 2020). Through community effort, at the time of writing in 2021 the template for methodology has been greatly refined and is

currently being integrated into a structured format making up the backend of a webapp (based on the model of COBIDAS Guidelines Checklist, see Table 1: Resources for GSP), which will help researchers document methodological details accurately in a simple, foolproof format.

To improve the ability of the MEEG community to achieve best practice in reporting, the ARTEM-IS Statement was drafted from the lessons of checklist adoption in aviation and surgery (Styles et al., 2021). The Statement is a call to action, describing how to: (1) improve clarity; (2) improve accuracy; (3) enhance documentation; and (4) deliver broad benefits. The Statement also recognizes that, to achieve this, (5) community effort will be necessary. That is to say, in order to improve the clarity of reporting, the scientific community needs tools that enable greater accuracy in reporting individual items, and this may be best achieved through enhanced documentation outside the traditional written summaries of journal articles. This approach would bring broad benefits to the community of MEEG stakeholders who aim to engage in replication or future meta-science, but will only be possible with the engagement of the broader MEEG community in drafting and clarifying the template, and creating tools of the greatest utility to users. The complete ARTEM-IS statement can be found in Table 1: Resources for GSP. Signatories to the statement are invited to participate in the further development of the ARTEM-IS template, and other similar initiatives in all fields of science are welcome to adopt the principles outlined in the ARTEM-IS Statement.

3.6.3. Data analysis and sharing: COBIDAS MEEG

As already noted, MEG and EEG communities have proactively championed GSP, with specialist society journals being the most common format for sharing them (e.g., (Donchin et al., 1977; Pivik et al., 1993; Picton et al., 2000; Duncan et al., 2009; Gross et al., 2013; Kane et al., 2017; Hari et al., 2018)). The Organization for Human Brain Mapping (OHBM) has developed GSP white papers in neuroimaging, which are collectively referred to as COBIDAS – an abbreviation for the “Committee on Best Practice in Data Analysis and Sharing”. There is a version for MRI-based methods (T. E. Nichols et al., 2017) and for MEEG (C. Pernet et al., 2018, 2020). In general, the COBIDAS Committee believes that guidelines should not be prescriptive, as data (pre)processing pipelines vary by analysis method. COBIDAS MEEG guidelines first discuss pitfalls for data acquisition, analysis, and sharing for resting state and task-related studies. However, the guidelines do include a set of Tables in the Appendix that list parameters that should be reported when preparing manuscripts and grants. Experimenters can pick and choose which of these tables are relevant at a given time. An electronic Python-based user-friendly checklist for MEG or EEG

studies supports this process. This is a complement to an already existing eCOBIDAS App for MRI-based data (see Table 1: Resources for GSP). Importantly, COBIDAS MEEG is consistent with other neurophysiological guidelines, e.g., International Federation for Clinical Neurophysiology (IFCN) (Kane et al., 2017), and uses BIDS-consistent terminology.

COBIDAS MEEG guidelines differ in 3 main ways from existing MEEG guidelines. First, they specifically include practices of reproducibility and data sharing. Second, they were prepared as a living document with 2 branches: (a) a WordPress blog for feedback and comments (see Table 1: Resources for GSP), which remains open for the next incarnation of the guidelines; (b) an Open Science Framework version-controlled White paper (C. Pernet et al., 2018). Third, the target population is much broader than for previous guidelines and now also includes not only neuroimagers/neurophysiologists, but also the hardware and software engineers and physicists who write MEEG papers, theses, grants, and prepare Registered Reports and clinical trials, as well as those who review and evaluate them.

Some problematic issues discussed in COBIDAS MEEG are: (i) omitted critical data acquisition details (see also (Šoškić et al., 2021)) from manuscripts and grants, e.g., low-pass or high-pass filtering, analog and digital (post-hoc) reference electrodes; (ii) the distortion issue if data are acquired with a physically-linked ear or mastoid reference; (iii) inadequate treatment of statistical power and related issues, e.g., minimal effect size estimation for features of interest, use of independent data from existing literature and/or pilot data for choosing regions/sensors of interest; (iv) protocols for rejecting artifactual trials, where trials need to be added to the original design to compensate for the loss of statistical power; (v) display items, e.g., figures, that do not adequately show variability measures, scales, or topography; (vi) inconsistent use of terminology – particularly pertinent for ERPs and canonical MEEG frequency bands, which may change over the lifespan, so if the actual frequency of the activity (and its regional distribution) is reported, confusion can be minimized; (vii) underspecified or omitted results of statistical analyses – including model assumptions, test statistics, effect sizes, and statistical maps for mass-univariate and multivariate analyses (Šoškić et al., 2021).

→ Technological and social contexts for MEEG continue to change, so ‘living documents’ are one solution to maintaining ever-changing guidelines for GSP.

→ The Brain Imaging Data Structure (BIDS) is a community-led standard for organizing, describing and sharing neuroimaging data becoming widely adopted in the MEEG community.

→ ARTEM-IS provides a template and tool for comprehensive reporting.

→ COBIDAS MEEG provides guidelines and checklists covering all parts of MEEG projects, specifically including reproducibility and data sharing.

4. Beyond the signal

Each scientific discipline has its own distinctive "way of doing things", and the more specialized a scientist becomes, the more they notice the devil in the details. Much more could be written about optimising experimental designs, data collection, signal processing, or any other specific aspect of the MEEG research cycle, but this is beyond the general scope of this article. In this section, we focus on social and ethical aspects of our discipline in particular, and science more generally.

As our branch of science deals directly with human beings, we have a duty of care for the duration that they are with us in the laboratory. Our concern for human well-being extends beyond participants and coworkers, including our own selves. The section 2.1. Human factors already discussed what that means for our experiments, but it also means, more generally, that we need to take care of our wider social communities. The remainder of this section will provide some perspectives on how a step back from scientifically focused details can improve our work, our field, and our society.

4.1. Social responsibilities and neuroethics

David Hume (Hume, 1739) famously argued for the distinction between "is" and "ought", separating factual judgements from moral ones, thus also reinforcing the division of scientific disciplines in "pure" and "applied" sciences (Proctor, 1991; Kincaid et al., 2007). There was, and still persists, an idea that “pure” sciences can be "value-free", or isolated from any moral considerations (Douglas, 2014). Irrespective of this larger philosophical debate, GSP requires an awareness of the moral implications and potential (mis)uses of scientific data, even when no immediate application is being investigated. Neuroscientific and particularly EEG-based research and clinical medicine is moving out of the lab into naturalistic real-world situations, including educational, home-based ambulatory settings (Makeig et al., 2009). With direct-to-consumer neurotechnology becoming increasingly available to the general public (Ienca et al., 2018), and with large international social media and tech companies

investing in neurotechnology (Moses et al., 2019; Musk & Neuralink, 2019), the possible ethical, legal, and societal implications of neuroscientific research are currently more pressing than ever.

BCIs, in particular, allow neural correlates of mental states to be identified and used in real time (Wolpaw & Wolpaw, 2012), and are a major focus of current consumer applications. With this technology, neuroprosthetics can be developed to support the motor-impaired (Wolpaw et al., 2002), and even the devices we use every day can be made to automatically adapt to our mental states (Zander & Kothe, 2011). However, it has also been shown that information can be obtained from brain activity that people did not intend to communicate, or did not wish to reveal (Schultze-Kraft et al., 2016; Zander et al., 2016). As such, neurotechnology may present a danger to the privacy of thought (Mecacci & Haselager, 2019). It is possible for a BCI-based device to actively attempt to manipulate a human participant's mental state, or to extract specific information from their brain activity, potentially without their knowledge or consent (Fairclough, 2017; Krol et al., 2020). This approach is, for example, the basis of the "guilty knowledge test", which attempts to identify incriminating response patterns in a criminal suspect's evoked brain activity (Rosenfeld et al., 2008). Currently, in some countries legal arguments can be made as to why a compelled test to analyse one’s brain activity does not (yet) violate any existing rights, such as those that protect against unreasonable searches or self-incrimination (Pardo & Patterson, 2013). However, what is legal (or not illegal) is not necessarily morally right, and regulatory law generally lags behind new scientific developments needing regulation. In the meantime, we should thus consider whether or in what way our work might be used to violate not just existing laws, but also any generic humane principles or human rights that do not yet exist, but ought to exist. The right to *mental integrity* is one such proposed right that MEEG researchers may inadvertently violate, or allow others to violate (Ienca & Andorno, 2017). Other issues related to neuroscience concern the potential of "brain hacking" (Ienca & Haselager, 2016), the influence of neurotechnology on our sense of agency (Haselager, 2013), and the societal impact of cognitive enhancement (Hyman, 2011), to name a few.

As such, neither neuroscience in general, nor MEEG research in particular, exists in moral isolation: indeed, neuroethics has a long history of thought on this topic. Scientists and researchers are encouraged to identify and discuss the potential ethical, legal, and societal implications of their research in their publications. In this vein, the Organization for Economic Co-operation and Development (OECD) recently adopted the Recommendation on Responsible Innovation in Neurotechnology (OECD, 2019), which contains nine principles aimed at guiding researchers to minimize societal risks. Among other things, it asks

researchers to consider the multidimensional societal implications that their work may have, to anticipate potential misuses, and to assess the safety of new neurotechnological developments. This process can be supported by, for example, Mecacci and Haselager's (Mecacci & Haselager, 2019) five criteria of accuracy, reliability, informativity, concealability, and enforceability, suggesting that we answer such questions as: Can this technology we are developing be used without the knowledge of the subject? Can it be used against their will? Here, the issues may not be obvious: even "classic" evoked potentials such as the P300 may be abused to extract information from the brain of unwitting participants (Martinovic et al., 2012) using methods such as cognitive probing (Krol et al., 2020).

Accepting that moral judgement of our work is unavoidable, we can proactively integrate ethical and societal considerations in our research and curricula, and look at neurotechnology not as a means (e.g., to gather data) but as an end: we can use the possibilities afforded by this technology to actively and explicitly "promote human flourishing" (Kellmeyer, 2018).

4.2. The map and the territory

> "*... In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters*" (Borges, 1946)

This text of Borges was used by Umberto Eco for an essay *On the impossibility of drawing a map of the empire on a scale of 1 to 1* (Eco et al., 1995). There Eco argues that, once folded, the map may not represent the folded map and may thus be unfaithful and useless. This reminds us of Heisenberg's uncertainty principle. Eco goes on to state that when the map is installed it does not represent the map itself over the territory. This is a sort of Russel's paradox applied to maps: a normal map cannot map itself. In other words: a) every 1:1 map always reproduces the territory unfaithfully; b) when the map is realized, the empire becomes unreproducible; and c) every 1:1 map of the empire decrees the end of the empire as such and therefore is the map of a territory that is not an empire.

If we replace the word *empire* with the word *brain*, we come closer to the topic of interest here, and recall that we do not study the brain, but different representations of the brain. Paraphrasing Eco, if recording all individual neurons is the brain's holy grail (*brah gra*), “it may decree the end of the scientific study of the brain as such and therefore a scientific field that is not concerned with the brain”, but with something else that is qualitatively different from it (e.g., individual neurons), and that misses its higher order emerging properties (e.g., the mind).

The aim of this literary and philosophical escapade is to bring us to examine notions of *orders of magnitude* in science. What we can observe for science in general, and neurosciences are no exception, is an increase in the number of almost everything, but the questions still remain mostly the same. For instance, while neuroscience moves from single case (or indeed single cell) studies to Big Data, with an amazing increase in all numbers (of brains, of bytes, of CPUs), do we truly understand the brain – not mentioning the mind – any better by just increasing numbers? Can we learn to ask different scientific questions from Big Data, to solve deeper questions in a way we have not done previously?

4.3. Rethinking our narratives: The challenge of “slow science”

4.3.1. The negative consequences of the culture of speed

As stated in previous sections, publication numbers in academia have exploded: more scientists are publishing more papers per capita than in the past (Bornmann & Mutz, 2015), while retracted papers are on the rise (Steen et al., 2013). For instance, more than 2 million papers are published every year, giving rise to the publishing paradox: the more we publish, the less we know (relative to the total number of publications). On top of representing a very lucrative business for publishers (Larivière et al., 2015), the increased number of publications and their cost tends to favour researchers and labs that have more funding and are located in Western countries, creating a *halo effect* around those labs and researchers (as well as other biases mentioned in section 2.1.3. Social factors influencing reasoning), reinforcing inequities, with no clear evidence of a benefit on the quality of science, and creating a detrimental lack of diversity (Fortin & Currie, 2013). This funding concentration results in decreasing funding availability, which gives rise to more competition and stronger incentives to be fast and productive to secure a position in academic research. The situation becomes althemore unacceptable that, in the US, less than 13% of PhD graduates currently succeed in securing long term positions in academia (Larson et al., 2014), so a number of

young scientists end up adapting their research agendas and contribute to creating a deleterious environment for a science they will not be able to contribute to in the long run.

Our current system, in which research is awarded in a way that favors this acceleration, shapes research practice dramatically. In particular, it favors a "culture of speed": incentivizing data accumulation with no time for reflection, fostering a competition that obstructs collaboration, imposing time wasted on applying for grants instead of actually doing science. It also has negative consequences for research quality, the heavy and blind use of metrics to grant funding or hiring encouraging abuse (Smaldino & McElreath, 2016) or strategic gaming (Chapman et al., 2019), decreasing trust in publications' quality (Vazire, 2017). Importantly, and despite being largely ignored, it also substantially impacts researchers' mental health, with increased levels of stress and frustration as a consequence of lack of time, work-overload, often coupled with fixed-term contracts, in uncaring work environments (see (Shaw & Ward, 2014) for an overview of this topic).

Finally, the culture of speed creates strong incentives against the GSP we have been presenting in this document. Planning projects, counteracting biases, promoting collaboration and inclusivity, writing good code (to name just a few) is key to GSP and takes time, and as we have explained in this document, failing to use GSP can lead to questionable research practices that have led to the replication crisis.

4.3.2. Main limitations of the open science movement's response to the culture of speed

The open science movement is currently the leading and most widely embraced response to the replication crisis. It promotes discussion, as well as different tools and platforms to facilitate the adoption of better research practices (reviewed in this paper and elsewhere, e.g., Renkewitz & Heene, 2019). It offers new incentives to encourage openness, integrity, and reproducibility, and the tools provide concrete solutions to current problems identified by the scientific community. For many, this is the new path to follow for generating trustworthy scientific output. However, as explained in the previous sub-section, research practices and structures have deeper problems that the open science movement – in spite of its good willingness – may be unable to fix. Without a deeper reflection on the roots of the culture of speed, it seems likely that newly created incentives may become tomorrow's new metrics, thus favouring "open science washing" practices or turning those practices into a list of boxes to tick to get one's work published or funded. As an example, the way in which the open science movement participates in the creation of new publication types (preprints, registered reports), contributing to the rise in publication numbers mentioned above,

currently remains largely unquestioned. Despite offering positive steps towards more transparency and better scientific practices, it fails to address the origins of the problem.

4.3.3. What do we do? A slow science perspective

A broader response may be offered by the slow science philosophy (Stengers, 2013). One core idea of this philosophy is to encourage a sustainable research praxis (Salo & Heikkinen, 2018). This includes emphasising the collective aspects of science rather than self-protecting our ideas, fostering reflection about our practices and the relevance of our research questions. By definition, slow science promotes more open science, where the goal is not to try to 'fix' but to rethink academia.

Importantly, slow science is not about forcing scientists to slow down their publication pace, even though some have explicitly suggested that this might be a desirable option (Frith, 2020). It is not about returning to some idealized golden age where scientists could focus deeply on their work, immune from the needs of society and productivity imperatives. Instead, slow science is about resisting the culture of speed and the idea of "wasting time" in research with futilities (Stengers, 2013). Unfortunately, overwork and multitasking have been traditionally celebrated, and work-life balance and mental health have not been taken seriously (Berg & Seeber, 2015). There is a need to question those narratives and how time pressure and time fragmentation (Ylijoki & Mäntylä, 2003) affects creativity and critical thinking on top of jeopardising wellbeing (Maestre, 2019). To foster those discussions and reflections, there has to be spaces where one can imagine and experiment with alternatives to the current system, even at a local level.

Putting those words into actions might seem unrealistic. However, initiatives already exist for paving the slow science way (see Table 1: Resources for GSP). Some examples are research frameworks that have emerged from group discussions at conferences: the DORA initiative (a.k.a. The Declaration of Research Assessment) promotes ways of evaluating researchers that go beyond single use of publication metrics, or the HIBAR Research Alliance, that aims to make research and innovation more integrated and aligned with society's critical problems. Other initiatives include discussion groups where scholars discuss and rethink academia and imagine a healthier research culture. Among them are such groups as Slow Science in Belgium or the Better Science initiative in Switzerland. Finally, some researchers have created independent institutes: the RONIN Institute that promotes research outside traditional institutions, or IGDORE, the Institute for Globally Distributed Open Research and Education, with a more integrated objective, that is,

improving the quality of science, science education and scientists' quality of life. Close to this philosophy, the Center for research and integrity (CRI) is exploring new ways of doing research, learning, or teaching while connecting with other parts of the society (industries, citizens, etc).

These existing networks, discussion groups, or more radical alternatives are all concrete examples that individual scientists can, at their level, promote healthier research practices. These examples should encourage scientists to explore other practices, question their relationship to academia, and start conversations with colleagues. For instance, the Café Culture Initiative promoted by the Wellcome Foundation offers a concrete way to do it. A discussion kit was developed to foster discussions about reimagining the research culture and working together. Some results about the research culture as experienced by researchers are already available on the Wellcome Foundation website.

In a nutshell, more needs to be done than just 'fixing academia' or improving scientific practice. Narratives, research frameworks, and practices need to be actively rethought. As summarized by (Lancaster et al., 2018), "*Making science better is not just about "creating better incentives", but a collective cultural shift beyond viewing competition and individualistic success as the sole defining feature of science.*"

Furthermore, academia is part of a larger ecosystem where every part must carry equal weight to keep a good balance. We need to emphasize our connections with the non-academic worlds and consider societies' needs and concerns.

4.3.4. GSP for environmental sustainability

Finally, there remains an important aspect of GSP that we have not discussed. Current scientific practices, like many human activities, are on an unsustainable long-term path. The current ecological collapse (The Intergovernmental Platform on Biodiversity and Ecosystem Services (IBPES) report, (Díaz et al., 2019)) and climate change (Intergovernmental Panel on Climate Change (IPCC) reports) is leading planet Earth to a state that may not allow exploiting the scientific knowledge created to its full extent. Part of the elaboration of the GSP for the future therefore relies on initiating immediate sustainable practices in all fields of science. To this end, many collective actions are taking place around the world. Over the past decade, most academic institutions and research foundations have created departments and initiatives dedicated to studying and acting for environmental sustainability on campuses and in research centers (see Table 1: Resources for GSP). These institutions,

along with academic grassroots collective initiatives are important local actors that could change GSP by setting local research agendas on more sustainable paths. As preliminary as these actions may be, they signal that the scientific community is progressively moving towards incorporating environmental considerations in the definition of GSP.

- → The culture of speed that currently defines academia creates wrong incentives that favour quantity over quality, competition over collaboration, leading to a stressful and uncaring work environment
- → The open science movement, although addressing some of those problems, fails to question the origins of this culture of speed and how to promote different ways of doing research
- → A slow science perspective can help to foster reflections on our current practice and the relevance of our research questions to rethink academica instead of merely fixing it. Some examples of existing initiatives are provided to illustrate what can be done.
- → Future GSP will have to take environmental aspects into account. Some actions and reflections are emerging locally but there is still a long way to go.

5. Conclusion

The 2020 LiveMEEG conference brought together international experts to discuss current progress and perspectives on good scientific practice in MEEG research. Virtual sessions and panels covered GSP across all stages of the research lifecycle, as well as broader topics that transcend experimental work, such as societal responsibilities and research culture. Some of these topics were examined in this manuscript. Strikingly, a common theme emerged regarding the value of collaborative work. Many contributors emphasized the benefits to be had from reaching out beyond our own office, lab, and institutional walls, and beyond our disciplines; from leveraging each other's competence, resources, and perspectives.

Indeed, there were many explicit suggestions of tools and methods for collaborative work: For example, collaborative meetings where scientists support each other to overcome human biases at early stages of experimental planning are proposed; pre-registration explicitly invites community feedback at early stages of the research process; different labs are encouraged to work together to reproduce findings and provide resources to the community; guides are being developed to cultivate further collaborative projects; and scientists across disciplines are reimagining science and academia in the current century.

Of course, not all work can be done collectively. Yet, even from those GSP perspectives that do not explicitly deal with shared work, a general theme of collaboration emerges – a common collaborative mindset, as it were, from which many different GSP guidelines emerge almost naturally. This mindset encourages scientists to see themselves not (primarily) as solitary experts in their specific field, but as part of a larger movement that ultimately shares the same goals of renewing truth and understanding. In essence, scientists are encouraged to either seek out collaborators where possible and appropriate, or otherwise imagine invisible or future collaborators who are eager to join or continue – but not usurp – the project.

The mere *idea* of having collaborators, imaginary or otherwise, almost automatically leads to practices that enable transparency, open science, and reproducibility. That is, assuming that the project at some point will be joined by or passed on to someone else provides clear, almost automatic incentives to uphold GSP: Collaborators must be able to access previous work, understand the full experimental philosophy, retrace analysis steps, and verify outcomes. As such, all efforts made towards reproducibility, be it by following known protocols, by providing annotated data, legible code, or full open resources, are in essence collaborative endeavours, anticipating potential collaborators. Even scientific publication and citation practices can be seen as a continuous collaboration across time and space, building upon previous work and creating new work for others to continue to build upon. Finally, future collaborators should, wherever possible, be saved from the disappointment of finding an avoidable flaw in the work on which they are building their own, emphasising the need for GSP also in the methodological details.

Now a key challenge for the shift towards a fully collaborative culture in MEEG is the need for coordination. Scientists have too long been trained to think and act individually, or in small mentor-advisee pairs, and generally lack training in cooperative action. Leading or contributing to a collective effort is in no way as easy and simple as conducting an individual project. This is perhaps where other areas of science and society can aid the MEEG community.

Thus, the collaborative mindset and traditional GSP are perfectly linked, but it also encourages scientists to “think big” in terms of developing new GSP, that requires collective efforts in open communities to develop worldwide standards, movements, and cultures. In summary, the many LiveMEEG 2020 contributors have shown that the search for "Eureka!", or "I have found it", is best envisioned as an endeavour to, instead, reach "*we* have found it" – "Eurékamen!"

6. Boxes

Box 1: Evolving GSP

GSP is under constant evolution. Current GSP may not be appropriate in the future, as new methods highlight limitations of current practice. Some examples are:

1. GSP for gamma-band oscillatory activity measurements has evolved dramatically after the discovery of a microsaccadic spike artifact in the EEG (Yuval-Greenberg et al., 2008). This spike artifact led to a serious reinterpretation of some early gamma-band oscillation studies, over a decade after these had been initially reported.

2. In the statistical domain, intuitive and widely used practices can become unacceptable after detailed critiques and scrutiny. Traditionally acceptable common practices in neuroimaging consisted of selecting regions of interest for further analysis with the same contrast that was used to report effect sizes in publications. This is now a well-known case of problematic circularity in data analysis (Kriegeskorte et al., 2009; Vul et al., 2009). Similarly, the failure to account for multiple comparisons in mass univariate data analysis (shown with irony by (Bennett et al., 2009), or by (T. Nichols & Hayasaka, 2003)) has become completely unacceptable.

3. Also regarding multiple comparisons, although results using the standard cluster-based approach (Maris and Oostenveld, 2007) seem to imply definite spatio-temporal limits, these are subject to an arbitrary threshold and should not be used as a way to localize effects in space and time (Sassenhagen & Draschkow, 2019). Therefore, the sensitivity of the cluster-based correction comes at the cost of being less specific when reporting results.

4. Similarly, increasing the number of factors in a statistical analysis (e.g., ANOVA) causes an exponential increase in the number of main effects and interactions being tested, which can dramatically increase the familywise Type I error rate (Cramer et al., 2016; Luck & Gaspelin, 2017). As a result, previous advice to include electrode/sensor as a factor in statistical analyses (Luck, 2005) has been replaced by advice to collapse across electrodes/sensors unless this factor is important for testing the scientific hypotheses (Luck, 2014).

Box 2: A brief historical perspective on GSP of MEEG analysis software

Prior to computers becoming tools for MEEG analysis in the later parts of the 20th century, many clinically based labs around the world performed "paper and pen" EEG data analysis.

Acquisition hardware used to print time-courses on either long rolls of graph paper roll or z-folding paper, and many researchers measured amplitudes directly on paper (especially in the clinic) and compared them to known amplitude square-wave calibration traces acquired at the beginning of the recording session. EEG frequencies of oscillatory activity could be read by counting the number of peaks within a nominal (1 second) time period, or by using a specially designed EEG ruler that allowed frequencies to be calculated from oscillatory waveform lengths. Experimental data were noted in tables and basic statistical analyses could be performed by hand, with reference to tables of critical values of the particular test statistic, which appeared as a function of degrees of freedom and threshold level of significance. During the same period, research labs also used digital computers, and a myriad of local toolboxes were created by scientists in labs around the world. Statistical analyses were often performed on mainframe computers using early versions of statistical software, such as SPSS, for example. A few toolboxes evolved and other new ones emerged and became widely used. They now enable complex data analysis of high-density MEEG datasets at a speed that users in the late 80s could not have dreamt of.

But with great power comes great responsibility, and with incentives pushing towards positive results, researchers often find themselves at best in situations of conflicts of interest, or at worst with opportunities for misconduct if the software is not carefully designed. When repeating tests with slightly different parameters requires just a few mouse clicks, it is easy to perform many different tests before completing the "final" to-be-reported analysis.

Therefore, a key tension for writing MEEG data analysis software has always been to, on the one hand, encapsulate data computations to make the whole process appear less complex than it actually is, while on the other hand make it difficult to make missteps, be it inadvertently or not. Leading MEEG toolboxes have different strategies to safeguard against most common errors. Below, we examine some widely used toolboxes for MEEG analysis (EEGLAB, FieldTrip, Brainstorm, MNE-python and SPM) and: (1) summarize how and where they were created, (2) identify reasons behind their different analysis approaches, (3) highlight excellence in specific domains, and (4) mention some of their limitations. All five toolboxes can run complete analysis pipelines, are Open Source (although 4 of the 5 do operate in the MATLAB closed source environment, which has an Open Source alternative called Octave). The most successful and widely used methods originally developed with one toolbox are usually ported and usable in the others. Advanced documentation and tutorials are available on their respective websites, and developers often run regional user training workshops.

Certainly one of the first, and currently the most cited open-source toolbox for EEG analysis is EEGLAB (Delorme & Makeig, 2004). This toolbox was created at the Institute of Neural Computation (now Swartz Center for Computational Neuroscience) at the University of California, San Diego, by Scott Makeig and Arnaud Delorme in the MATLAB scientific programming language (support for Octave starting in 2021). Makeig brought independent component analysis (ICA) to EEG data analysis (Makeig et al., 1997), and has since continuously promoted the use of this blind source separation method for both artifact rejection and data analysis (see also (Martínez-Cancino et al., 2021)). This history makes EEGLAB more suited for EEG analysis (not so much for MEG, although MEG file formats can be read). Whereas other toolboxes mostly use ICA for artifact correction only, EEGLAB was designed to also incorporate independent components as virtual channels and perform standard analyses with these, as well as specific analyses such as component clustering, or network analysis. As a strategy to guide users and promote GSP, this toolbox uses a simple graphical user interface, with sequential processing steps generally arranged along specific menus, which makes it particularly well suited for experimenters with little or no programming experience. Users of EEGLAB can therefore easily create scripts for most point-and-click commands, which can be further edited. EEGLAB also forms the centerpiece for a large and constantly growing plug-in (a.k.a. Extensions) ecosystem that allows incorporation of cutting-edge methods, procedures, and pipelines in EEGLAB workflows. Current developments of this toolbox include mobile EEG (Ojeda et al., 2014), general linear modeling of EEG activity (Delorme et al., 2021), and interfacing with online computing services (Martínez-Cancino et al., 2021).

The FieldTrip toolbox has been developed since 2003 by a team led by Robert Oostenveld (Oostenveld et al., 2011) at the Donders Center for Cognitive Neuroimaging, now the Donders Institute for Brain, Cognition and Behaviour at the Radboud University, Nijmegen, the Netherlands. This complete MEEG toolbox is particularly well suited to oscillatory data analysis and source modelling. Its so-called cluster-based permutation statistical analysis framework (Maris & Oostenveld, 2007) is the current gold-standard method to correct for multiple comparisons in MEEG data analysis. A unique feature of this toolbox is its ability to handle trials of varying duration and it is also particularly well adapted for human intracranial and invasive animal electrophysiology. This toolbox promotes GSP with a well organized command line interface where configuration parameters are systematically separated from data. This separation promotes a robust structurization of analysis scripts, promotes reproducibility, and makes it suitable for experimenters with basic MATLAB experience, as well as more advanced users. Current efforts towards automated generation of reproducible pipelines (van Es et al. submitted current issue) will further facilitate this strategy.

Brainstorm has a very sophisticated graphical user interface. This toolbox was originally created in MATLAB by Sylvain Baillet in the late 1990s (Baillet et al., 2000) at the Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale (LENA) in Paris. It has been entirely revamped, further developed, and professionally managed by François Tadel since 2008 in Paris, and in Sylvain Baillet's team at the Montreal Neurological Institute at McGill University in Montreal (Tadel et al., 2011, 2019), with large contributions from John Mosher (University of Texas Health Science Center at Houston) and Richard Leahy (University of Southern California). It enforces GSP with powerful data exploration and interactive visualization tools, and a database engine that handles data files transparently for the user. It is therefore particularly well suited to adopt the current move towards standardized data handling (see BIDS section) (Niso et al., 2019). Importantly, Brainstorm can handle multiple data types in addition to MEG, scalp and intracranial EEG within the same protocol. These include fNIRS and multiunit electrophysiology. Brainstorm also includes an advanced pipeline generator and a modular "process" ecosystem that allows incorporating virtually any analysis method into its environment. Its robust software architecture also makes it a choice platform environment to handle data across several of the toolboxes presented here, e.g., with calls to FieldTrip and MNE-python from within Brainstorm.

The MNE-Python toolbox is the most recent addition to this toolbox list. It is based on a Python reimplementation (Gramfort et al., 2013) of the original MNE-C Minimum Norm Estimate (MNE) algorithm written in C (Gramfort et al., 2014; Hämäläinen & Ilmoniemi, 1994). Alexandre Gramfort continues to lead the development of this toolbox that has become the cornerstone of a robust ecosystem with powerful analysis and visualization tools, and a large and lively contributor community organized with modern development tools. The toolbox can handle MEG, EEG, iEEG (ECoG and DBS), EKG, EMG, fNIRS and other data types. MNE-Python and its interface with the Scikit-learn Python toolkit for machine learning (Pedregosa et al. 2011) make it a choice toolbox for decoding and machine learning analysis of MEEG signals. Furthermore, it implements a comprehensive list of methods for source imaging ranging from dipole fitting, linear methods (dSPM, sLORETA), beamformers, (RAP-)MUSIC and more recent nonlinear approaches. GSP is enforced in this environment in particular through the object-oriented interface that ensures that specific processes can only be applied to designated appropriate data objects, as well as with the development of child projects such as MNE-BIDS (Appelhoff et al., 2019), to facilitate BIDS integration. The rich and open-source Python ecosystem and the strong bonds of the original developers of the Freesurfer (Dale & Sereno, 1993; Dale et al., 1999;

Fischl, 2012) MRI segmentation toolbox make it easy to interface with those widely used tools.

Finally, SPM was originally developed for the analysis of positron emission tomography and subsequently functional MRI data by Karl Friston in London (Friston et al., 2007). SPM also allows to analyze MEEG data (Litvak et al., 2011) using the general linear model (GLM) and a Bayesian framework for source reconstruction. SPM also supports analyses of effective connectivity and fitting biophysically realistic neural mass models to MEEG data using dynamic causal modelling (DCM) framework (Kiebel et al., 2008). In its current incarnation, SPM handles fMRI, PET, EEG, and MEG data in a MATLAB environment, with an emphasis on scripting analyses for maintaining transparent data processing pipelines.

As already mentioned, these toolboxes have different origins which explain their design choices, current orientations, implemented features, and strengths. But they all provide robust building blocks to become the home of future cutting-edge methodologies.

7. Tables

Table 1: Resources for Good Scientific Practice in MEEG research

Table 2: Widely used MEEG analysis toolboxes

8. Author Contributions

Author contributions are reported following the CRediT Contributor Roles:
Conceptualization: Guiomar Niso, Laurens R. Krol, Anne-Sophie Dubarry, Clément François, Manuel Mercier, and Maximilien Chaumon.
Project administration: Guiomar Niso and Maximilien Chaumon.
Supervision: Guiomar Niso, Laurens R. Krol, and Maximilien Chaumon.
Writing - original draft: Guiomar Niso, Laurens R. Krol, Etienne Combrisson, Anne-Sophie Dubarry, Madison A. Elliott, Yseult Héjja-Brichard, Sophie K. Herbst, Karim Jerbi, Vanja Kovic, Katia Lehongre, Steven J. Luck, Manuel Mercier, John C. Mosher, Yuri G. Pavlov, Aina Puce, Antonio Schettino, Daniele Schön, Walter Sinnott-Armstrong, Bertille Somon, Anđela Šoškić, Suzy J. Styles, Roni Tibon, Martina G. Vilas, Marijn van Vliet, and Maximilien Chaumon.

Writing - review & editing: Guiomar Niso, Laurens R. Krol, Etienne Combrisson, Anne-Sophie Dubarry, Madison A. Elliott, Yseult Héjja-Brichard, Sophie K. Herbst, Karim Jerbi, Vanja Kovic, Katia Lehongre, Steven J. Luck, Manuel Mercier, John C. Mosher, Yuri G. Pavlov, Aina Puce, Antonio Schettino, Daniele Schön, Walter Sinnott-Armstrong, Bertille Somon, Anđela Šoškić, Suzy J. Styles, Roni Tibon, Martina G. Vilas, Marijn van Vliet, and Maximilien Chaumon.

9. Funding

GN was supported by Spanish Ministry of Science and Innovation through grant PID2020-119458RJ-I00 and by the AXA Research Fund; MAE was supported by UBC 4YF; SJL was supported by National Institute of Mental Health (R25MH080794 and R01MH087450); JCM was supported by National Institutes of Health (NIH) (NIBIB); YGP was supported by Russian Foundation for Basic Research (RFBR) #1901300027; AP, GN and MS are supported by an NIH CRCNS: US-France Data Sharing Proposal (NIBIB (USA) R01 EB030896 and ANR-20-NEUC-0004-01); AS is employed as Coordinator Open Science at Erasmus Research Services (Erasmus University Rotterdam); WSA was supported by John Templeton Foundation; BS was supported by The Artificial and Natural Intelligence Toulouse Institute; RT was supported by British Academy Postdoctoral Fellowship (SUAI/028 RG94188 pf170046); MC was supported by Program Investissements davenir ANR - 10 - IAIHU - 06.

10. Conflict of Interest

The authors declare no conflict of interest related to this work.

11. References

Abelson, R. P. (1995). *Statistics As Principled Argument* (1st edition). Psychology Press.

Amer, T. S., & Ravindran, S. (2010). The Effect of Visual Illusions on the Graphical Display of Information. *Journal of Information Systems*, *24*(1), 23–42. https://doi.org/10.2308/jis.2010.24.1.23

Appelhoff, S., Sanderson, M., Brooks, T., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A.,

Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896

Babbage, C. (1957, November 10). Work with new electronic 'brains' opens field for army math experts. *The Times*, 65.

Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7–7. https://doi.org/10.1167/14.4.7

Bagic, A. I., Knowlton, R. C., Rose, D. F., Ebersole, J. S., & Committee, A. C. P. G. (CPG). (2011). American clinical magnetoencephalography society clinical practice guideline 1: Recording and analysis of spontaneous cerebral activity. *Journal of Clinical Neurophysiology*, *28*(4), 348–354.

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, *18*(6), 14–30. https://doi.org/10.1109/79.962275

Baillet, S., Mosher, J., & Leahy, R. (2000). BrainStorm beta release: A Matlab software package for MEG signal processing and source localization and visualization. *Neuroimage*, *11*. https://doi.org/10.1016/S1053-8119(00)91843-3

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2020). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*.

Baron, J. (2000). *Thinking and Deciding*. Cambridge University Press.

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, *4*(4), 417–422. https://doi.org/10.1093/scan/nsp053

Berg, M., & Seeber, B. (2015). *The slow professor: Challenging the culture of speed in the academy.* University of Toronto Press.

Beveridge, W. I. (2004). *The Art of Scientific Investigation*. The Blackburn Press.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Bolt, T., Nomi, J. S., Bzdok, D., & Uddin, L. Q. (2021). Educating the future generation of researchers: A cross-disciplinary survey of trends in analysis methods. *PLOS Biology*, *19*(7), e3001313. https://doi.org/10.1371/journal.pbio.3001313

Borges, J. L. (1946, March). Del rigor en la ciencia. *Los Anales de Buenos Aires*.

Borland, D., & Taylor Ii, R. M. (2007). Rainbow Color Map (Still) Considered Harmful. *IEEE Computer Graphics and Applications*, *27*(2), 14–17. https://doi.org/10.1109/MCG.2007.323435

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis

based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222. https://doi.org/10.1002/asi.23329

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Brouwers, K., Cooke, A., Chambers, C. D., Henson, R., & Tibon, R. (2020). Evidence for prereg posters as a platform for preregistration. *Nature Human Behaviour*, *4*(9), 884–886. https://doi.org/10.1038/s41562-020-0868-z

Buccino, A. P., Hurwitz, C. L., Garcia, S., Magland, J., Siegle, J. H., Hurwitz, R., & Hennig, M. H. (2020). SpikeInterface, a unified framework for spike sorting. *ELife*, *9*, e61834. https://doi.org/10.7554/eLife.61834

Bujack, R., Turton, T. L., Samsel, F., Ware, C., Rogers, D. H., & Ahrens, J. (2018). The Good, the Bad, and the Ugly: A Theoretical Framework for the Assessment of Continuous Colormaps. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 923–933. https://doi.org/10.1109/TVCG.2017.2743978

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

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature Reviews. Neuroscience*, *13*(6), 407–420. https://doi.org/10.1038/nrn3241

Casadevall, A., Steen, R. G., & Fang, F. C. (2014). Sources of error in the retracted scientific literature. *The FASEB Journal*, *28*(9), 3847–3855. https://doi.org/10.1096/fj.14-256735

Chambers, C. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C., & Tzavella, L. (2020). *The past, present, and future of Registered Reports* [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/43298

Chambers, J. (2018). *Graphical Methods for Data Analysis*. CRC Press.

Chapman, C. A., Bicca-Marques, J. C., Calvignac-Spencer, S., Fan, P., Fashing, P. J., Gogarten, J., Guo, S., Hemingway, C. A., Leendertz, F., Li, B., Matsuda, I., Hou, R., Serio-Silva, J. C., & Chr. Stenseth, N. (2019). Games academics play and their consequences: How authorship, h-index and journal impact factors are shaping the future of academia. *Proceedings of the Royal Society B: Biological Sciences*,

286(1916), 20192047. https://doi.org/10.1098/rspb.2019.2047

Choy, T., Baker, E., & Stavropoulos, K. (2021). Systemic Racism in EEG Research: Considerations and Potential Solutions. *Affective Science*. https://doi.org/10.1007/s42761-021-00050-0

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

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). https://doi.org/10.1111/psyp.13437

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT Press. https://books.google.fr/books?hl=fr&lr=&id=rDKkAgAAQBAJ&oi=fnd&pg=PR5&dq=analysing+neural+time+series+data+cohen&ots=g85iy-41y_&sig=NiVVTEHesbjMpA6u-EQwK4lr_a4

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

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, *250*, 126–136. https://doi.org/10.1016/j.jneumeth.2015.01.010

Community, T. T. W., Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O'Reilly, M., & Whitaker, K. (2019). *The Turing Way: A Handbook for Reproducible Data Science*. Zenodo. https://doi.org/10.5281/zenodo.3233986

Coren, S., & Girgus, J. S. (2020). *Seeing is deceiving: The psychology of visual illusions*. Routledge.

Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., Waldorp, L. J., & Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647. https://doi.org/10.3758/s13423-015-0913-5

Dalal, S. S., Baillet, S., Adam, C., Ducorps, A., Schwartz, D., Jerbi, K., Bertrand, O., Garnero, L., Martinerie, J., & Lachaux, J.-P. (2009). Simultaneous MEG and intracranial EEG recordings during attentive reading. *NeuroImage*, *45*(4), 1289–1304. https://doi.org/10.1016/j.neuroimage.2009.01.017

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Dale, A. M., & Sereno, M. I. (1993). Improved Localizadon of Cortical Activity by Combining

EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience*, *5*(2), 162–176. https://doi.org/10.1162/jocn.1993.5.2.162

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Truong, D., Martinez-Cancino, R., Pernet, C., Sivagnanam, S., Yoshimoto, K., Poldrack, R., Majumdar, A., & Makeig, S. (2021). Tools for Importing and Evaluating BIDS-EEG Formatted Data. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 210–213. https://doi.org/10.1109/NER49283.2021.9441399

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Ozge Buzbas, E. (2020). The case for formal methodology in scientific reform. *Royal Society Open Science*, *8*(3), 200805. https://doi.org/10.1098/rsos.200805

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471). https://doi.org/10.1126/science.aax3100

Donchin, E., Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., Hillyard, S. A., & Sutton, S. (1977). Publication criteria for studies of evoked potentials (EP) in man: Methodology and publication criteria. In *Progress in clinical neurophysiology: Vol. 1. Attention, voluntary contraction and event-related cerebral potentials* (pp. 1–11). Karger.

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., Knight, R. T., Shestyuk, A., & Voytek, B. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665. https://doi.org/10.1038/s41593-020-00744-x

Douglas, T. (2014). The dual-use problem, scientific isolationism and the division of moral labour. *Monash Bioethics Review*, *32*(1–2), 86–105. https://doi.org/10.1007/s40592-014-0004-9

Dubarry, A.-S., Badier, J.-M., Trébuchon-Da Fonseca, A., Gavaret, M., Carron, R., Bartolomei, F., Liégeois-Chauvel, C., Régis, J., Chauvel, P., Alario, F.-X., & Bénar, C.-G. (2014). Simultaneous recording of MEG, EEG and intracerebral EEG during visual stimulation: From feasibility to single-trial analysis. *NeuroImage*, *99*, 548–558. https://doi.org/10.1016/j.neuroimage.2014.05.055

Ducorps, A., Hugueville, L., Lemarechal, J.-D., Schwartz, D., & Yahia-Cherif, L. (2010, March 28). A user-friendly software suite for MEG, EEG, EcoG / Intracranial-EEG data analysis. *Frontiers in Neuroscience*. Biomag, 17th International Conference on Biomagnetism, Dubrovnik, Croatia. https://doi.org/10.3389/conf.fnins.2010.06.00107

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045

Eco, U., Sterling, D., & Weaver, W. (1995). *How to Travel with a Salmon & Other Essays* (First edition). Mariner Books.

Es, M. W. J. van, Spaak, E., Schoffelen, J.-M., & Oostenveld, R. (2021). Reducing the efforts to create reproducible analysis code with FieldTrip. *BioRxiv*, 2021.02.05.429886. https://doi.org/10.1101/2021.02.05.429886

Etienne, A., Laroia, T., Weigle, H., Afelin, A., Kelly, S. K., Krishnan, A., & Grover, P. (2020). Novel Electrodes for Reliable EEG Recordings on Coarse and Curly Hair. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2020*, 6151–6154. https://doi.org/10.1109/EMBC44109.2020.9176067

Fahimi Hnazaee, M., Wittevrongel, B., Khachatryan, E., Libert, A., Carrette, E., Dauwe, I., Meurs, A., Boon, P., Van Roost, D., & Van Hulle, M. M. (2020). Localization of deep brain activity with scalp and subdural EEG. *NeuroImage*, *223*, 117344. https://doi.org/10.1016/j.neuroimage.2020.117344

Fairclough, S. H. (2017). Physiological Computing and Intelligent Adaptation. In M. Jeon (Ed.), *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 539–556). Academic Press. https://doi.org/10.1016/B978-0-12-801851-4.00020-3

Farrens, J. L., Simmons, A. M., Luck, S. J., & Kappenman, E. S. (2020). *Electroencephalogram (EEG) Recording Protocol for Cognitive and Affective Human Neuroscience Research* [Preprint]. Protocol Exchange. https://doi.org/10.21203/rs.2.18328/v3

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fortin, J., & Currie, D. (2013). Big Science vs. Little Science: How Scientific Impact Scales with Funding. *Undefined*. https://www.semanticscholar.org/paper/Big-Science-vs.-Little-Science%3A-How-Scientific-with-Fortin-Currie/9334c460b5b19e1d0d607084750ece6462ef2e19

Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (Eds.). (2007).

Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press. http://store.elsevier.com/product.jsp?isbn=9780123725608

Frith, U. (2020). Fast Lane to Slow Science. *Trends in Cognitive Sciences*, *24*(1), 1–2. https://doi.org/10.1016/j.tics.2019.10.007

From raw MEG/EEG to publication: How to perform MEG/EEG group analysis with free academic software. | *Frontiers Research Topic*. (2018). https://www.frontiersin.org/research-topics/5158/

Gau, R., Gould van Praag, C., van Mourik, T., Wiebels, K., Adolfi, F., Scarpazza, C., Ruotsalainen, I., Tepper, A., Sjoerds, Z., Simon, J., Klapwijk, E., Hortensius, R., Bartlett, J., & Moreau, D. (2019). *COBIDAS checklist*. https://doi.org/10.17605/OSF.IO/ANVQY

Gavaret, M., Dubarry, A.-S., Carron, R., Bartolomei, F., Trébuchon, A., & Bénar, C.-G. (2016). Simultaneous SEEG-MEG-EEG recordings Overcome the SEEG limited spatial sampling. *Epilepsy Research*, *128*, 68–72. https://doi.org/10.1016/j.eplepsyres.2016.10.013

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*. Department of statistics, Columbia University. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, *3*(1), 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., Churchill, N. W., Cohen, A. L., Craddock, R. C., Devenyi, G. A., Eklund, A., Esteban, O., Flandin, G., Ghosh, S. S., Guntupalli, J. S., Jenkinson, M., Keshavan, A., Kiar, G., Liem, F., … Poldrack, R. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Computational Biology*, *13*(3), e1005209. https://doi.org/10.1371/journal.pcbi.1005209

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*, 160044. https://doi.org/10.1038/sdata.2016.44

Gorgolewski, K. J., & Poldrack, R. A. (2016). A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research. *PLOS Biology*, *14*(7), e1002506. https://doi.org/10.1371/journal.pbio.1002506

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & H??m??l??inen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(7 DEC), 1–13. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & H??m??l??inen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Gregory, R. L. (1968). Visual illusions. *Scientific American*, *219*(5), 66–79.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J. R., van Wassenhove, V., Wibral, M., & Schoffelen, J. M. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363. https://doi.org/10.1016/j.neuroimage.2012.10.001

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*(1), 35–42. https://doi.org/10.1007/BF02512476

Handy, T. C. (Ed.). (2005). *Event-related potentials: A methods handbook*. MIT Press.

Hari, R., Baillet, S., Barnes, G., Burgess, R., Forss, N., Gross, J., Hämäläinen, M., Jensen, O., Kakigi, R., Mauguière, F., Nakasato, N., Puce, A., Romani, G.-L., Schnitzler, A., & Taulu, S. (2018). IFCN-endorsed practical guidelines for clinical magnetoencephalography (MEG). *Clinical Neurophysiology*, *129*(8), 1720–1747. https://doi.org/10.1016/j.clinph.2018.03.042

Hari, R., & Puce, A. (2017). *MEG-EEG Primer* (1st ed.). Oxford University Press. https://doi.org/10.1093/med/9780190497774.001.0001

Hartmann, K. G., Schirrmeister, R. T., & Ball, T. (2018). EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *ArXiv:1806.01875 [Cs, Eess, q-Bio, Stat]*. http://arxiv.org/abs/1806.01875

Haselager, P. (2013). Did I Do That? Brain-Computer Interfacing and the Sense of Agency. *Minds and Machines*, *23*(3), 405–418. https://doi.org/10.1007/s11023-012-9298-7

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Hayasaka, S., & Nichols, T. E. (2003). Validating cluster size inference: Random field and permutation methods. *NeuroImage*, *20*(4), 2343–2356. https://doi.org/10.1016/j.neuroimage.2003.08.003

Hedderich, D. M., & Eickhoff, S. B. (2021). Machine learning for psychiatry: Getting doctors at the black box? *Molecular Psychiatry*, *26*(1), 23–25. https://doi.org/10.1038/s41380-020-00931-z

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29.

Hinss, M. F., Somon, B., Dehais, F., & Roy, R. N. (2021). Open EEG Datasets for Passive Brain-Computer Interface Applications: Lacks and Perspectives. *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 686–689. https://doi.org/10.1109/NER49283.2021.9441214

Hintze, J. L., & Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, *52*(2), 181–184. https://doi.org/10.1080/00031305.1998.10480559

Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A.-L. (2020). *The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines*. MetaArXiv. https://doi.org/10.31222/osf.io/afb9p

Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B. L., Gorgolewski, K. J., Groen, I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C. J., Jas, M., Knight, R., Lachaux, J.-P., … Hermes, D. (2019). IEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. *Scientific Data*, *6*(1), 102. https://doi.org/10.1038/s41597-019-0105-7

Hume, D. (1739). *A Treatise of Human Nature*. Printed for John Noon.

Hyman, S. E. (2011). Cognitive enhancement: Promises and perils. *Neuron*, *69*(4), 595—598. https://doi.org/10.1016/j.neuron.2011.02.012

Ienca, M., & Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sciences, Society and Policy*, *13*(1), 5. https://doi.org/10.1186/s40504-017-0050-1

Ienca, M., & Haselager, P. (2016). Hacking the brain: Brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology*, *18*(2), 117–129. https://doi.org/10.1007/s10676-016-9398-9

Ienca, M., Haselager, P., & Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nature Biotechnology*, *36*(9), 805–810. https://doi.org/10.1038/nbt.4240

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS*

Medicine, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jaiswal, A., Nenonen, J., Stenroos, M., Gramfort, A., Dalal, S. S., Westner, B. U., Litvak, V., Mosher, J. C., Schoffelen, J.-M., Witton, C., Oostenveld, R., & Parkkonen, L. (2020). Comparison of beamformer implementations for MEG source localization. *NeuroImage*, *216*, 116797. https://doi.org/10.1016/j.neuroimage.2020.116797

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st edition). Farrar, Straus and Giroux.

Kakisaka, Y., Kubota, Y., Wang, Z. I., Piao, Z., Mosher, J. C., Gonzalez-Martinez, J., Jin, K., Alexopoulos, A. V., & Burgess, R. C. (2012). Use of simultaneous depth and MEG recording may provide complementary information regarding the epileptogenic region. *Epileptic Disorders*, *14*(3), 298–303. https://doi.org/10.1684/epd.2012.0517

Kane, N., Acharya, J., Beniczky, S., Caboclo, L., Finnigan, S., Kaplan, P. W., Shibasaki, H., Pressler, R., & van Putten, M. J. A. M. (2017). A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. *Clinical Neurophysiology Practice*, *2*, 170–185. https://doi.org/10.1016/j.cnp.2017.07.002

Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, *225*, 117465.

Kappenman, E. S., Farrens, J., Zhang, W., Stewart, A. X., & Luck, S. J. (2020). *ERP CORE*. https://osf.io/thsqg/

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2010.01009.x

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/10.1016/j.bpsc.2015.11.007

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., & Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography: Guidelines for EEG and MEG. *Psychophysiology*, *51*(1), 1–21. https://doi.org/10.1111/psyp.12147

Kellmeyer, P. (2018). Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices. *Neuroethics*. https://doi.org/10.1007/s12152-018-9371-x

Kiebel, S. J., Garrido, M. I., Moran, R. J., & Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics*, *2*(2), 121–136. https://doi.org/10.1007/s11571-008-9038-0

Kincaid, H., Dupré, J., & Wylie, A. (2007). *Value-Free Science?: Ideals and Illusions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195308969.001.0001

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2014.01.002

Kirchberger, K., Hummel, C., & Stefan, H. (1998). Postoperative multichannel magnetoencephalography in patients with recurrent seizures after epilepsy surgery. *Acta Neurologica Scandinavica*, *98*(1), 1–7. https://doi.org/10.1111/j.1600-0404.1998.tb07370.x

Klein, G. (2007). Performing a project premortem. *Harvard Business Review*, *85*(9), 18–19.

Koessler, L., Cecchin, T., Colnat-Coulbois, S., Vignal, J.-P., Jonas, J., Vespignani, H., Ramantani, G., & Maillard, L. G. (2015). Catching the Invisible: Mesial Temporal Source Contribution to Simultaneous EEG and SEEG Recordings. *Brain Topography*, *28*(1), 5–20. https://doi.org/10.1007/s10548-014-0417-z

Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In J. S. Gero & F. Sudweeks (Eds.), *Artificial Intelligence in Design '96* (pp. 151–170). Springer Netherlands. https://doi.org/10.1007/978-94-009-0279-4_9

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Krol, L. R., Haselager, P., & Zander, T. O. (2020). Cognitive and affective probing: A tutorial and review of active learning for neuroadaptive technology. *Journal of Neural Engineering*, *17*(1), 012001. https://doi.org/10.1088/1741-2552/ab5bb5

Krol, L. R., Pawlitzki, J., Lotte, F., Gramann, K., & Zander, T. O. (2018). SEREEGA: Simulating event-related EEG activity. *Journal of Neuroscience Methods*, *309*, 13–24. https://doi.org/10.1016/j.jneumeth.2018.08.001

Lakens, D. (2021). *Sample Size Justification*. PsyArXiv. https://doi.org/10.31234/osf.io/9d3yf

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Lancaster, A. K., Thessen, A. E., & Virapongse, A. (2018). A new paradigm for the scientific enterprise: Nurturing the ecosystem. *F1000Research*, *7*, 803. https://doi.org/10.12688/f1000research.15078.1

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, *10*(6), e0127502. https://doi.org/10.1371/journal.pone.0127502

Larson, R. C., Ghaffarzadegan, N., & Xue, Y. (2014). Too Many PhD Graduates or Too Few Academic Job Openings: The Basic Reproductive Number R0 in Academia. *Systems Research and Behavioral Science*, *31*(6), 745–750. https://doi.org/10.1002/sres.2210

Larson-Prior, L. J., Oostenveld, R., Della Penna, S., Michalareas, G., Prior, F., Babajani-Feremi, A., Schoffelen, J. M., Marzetti, L., de Pasquale, F., Di Pompeo, F., Stout, J., Woolrich, M., Luo, Q., Bucholz, R., Fries, P., Pizzella, V., Romani, G. L., Corbetta, M., & Snyder, A. Z. (2013). Adding dynamics to the Human Connectome Project with MEG. *NeuroImage*, *80*, 190–201. https://doi.org/10.1016/j.neuroimage.2013.05.056

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004

Levitis, E., van Praag, C. D. G., Gau, R., Heunis, S., DuPre, E., Kiar, G., Bottenhorn, K. L., Glatard, T., Nikolaidis, A., Whitaker, K. J., Mancini, M., Niso, G., Afyouni, S., Alonso-Ortiz, E., Appelhoff, S., Arnatkeviciute, A., Atay, S. M., Auer, T., Baracchini, G., … Maumet, C. (2021). Centering inclusivity in the design of online conferences—An OHBM–Open Science perspective. *GigaScience*, *10*(8). https://doi.org/10.1093/gigascience/giab051

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain: A Journal of Neurology*, *106 (Pt 3)*, 623–642. https://doi.org/10.1093/brain/106.3.623

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

Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W. Y., Kidd, C., Marcelle, E., … Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron*, *109*(13), 2047–2074. https://doi.org/10.1016/j.neuron.2021.06.002

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition). The MIT Press.

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

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*, e13793.

Luo, Y., & Zhao, J. (2019). Motivated Attention in Climate Change Perception and Action. *Frontiers in Psychology*, *10*, 1541. https://doi.org/10.3389/fpsyg.2019.01541

Maestre, F. T. (2019). Ten simple rules towards healthier research labs. *PLOS Computational Biology*, *15*(4), e1006914. https://doi.org/10.1371/journal.pcbi.1006914

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, *73*(2), 95–100. https://doi.org/10.1016/j.ijpsycho.2008.11.008

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, *94*(20), 10979–10984.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncalves, M., Jwa, A., & Poldrack, R. A. (2021). *OpenNeuro: An open resource for sharing of neuroimaging data* (p. 2021.06.28.450168). https://www.biorxiv.org/content/10.1101/2021.06.28.450168v2

Martínez-Cancino, R., Delorme, A., Truong, D., Artoni, F., Kreutz-Delgado, K., Sivagnanam, S., Yoshimoto, K., Majumdar, A., & Makeig, S. (2021). The open EEGLAB portal Interface: High-Performance computing with EEGLAB. *NeuroImage*, *224*, 116778. https://doi.org/10.1016/j.neuroimage.2020.116778

Martinovic, I., Davies, D., Frank, M., Perito, D., Ros, T., & Song, D. (2012). On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces. *21st USENIX Security Symposium (USENIX Security 12)*, 143–158.

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2271–2279. https://doi.org/10.1080/17470218.2012.711335

Mecacci, G., & Haselager, P. (2019). Identifying Criteria for the Evaluation of the Implications of Brain Reading for Mental Privacy. *Science and Engineering Ethics*, *25*(2), 443–461. https://doi.org/10.1007/s11948-017-0003-3

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

Miller, G. (2006). Scientific publishing. A scientist's nightmare: Software problem leads to five retractions. *Science (New York, N.Y.)*, *314*(5807), 1856–1857. https://doi.org/10.1126/science.314.5807.1856

Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*, *10*(1), 3096. https://doi.org/10.1038/s41467-019-10994-4

Mosher, J. C., & Funke, M. E. (2020). Towards Best Practices in Clinical Magnetoencephalography: Patient Preparation and Data Acquisition. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, *37*(6), 498–507. https://doi.org/10.1097/WNP.0000000000000542

Munzner, T. (2014). *Visualization Analysis and Design*. A K Peters/CRC Press. https://doi.org/10.1201/b17511

Musk, E. & Neuralink. (2019). An integrated brain-machine interface platform with thousands of channels. *BioRxiv*. https://doi.org/10.1101/703801

Nave, K., Hannon, E., & Snyder, J. S. (2020, September 23). *Replication and Extension of Nozaradan, Peretz, Missal and Mouraux (2011)*. osf.io/rpvde

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M. P., Poldrack, R. A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D. C., White, T., & Yeo, B. T. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3), 299–303. https://doi.org/10.1038/nn.4500

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. https://doi.org/10.1191/0962280203sm341ra

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

Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., Henson, R. N., Jas, M., Litvak, V., T. Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F.,

Wexler, J., & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, *5*(1), 180110. https://doi.org/10.1038/sdata.2018.110

Niso, G., Rogers, C., Moreau, J. T., Chen, L. Y., Madjar, C., Das, S., Bock, E., Tadel, F., Evans, A. C., Jolicoeur, P., & Baillet, S. (2016). OMEGA: The Open MEG Archive. *NeuroImage*, *124*, 1182–1187. https://doi.org/10.1016/j.neuroimage.2015.04.028

Niso, G., Tadel, F., Bock, E., Cousineau, M., Santos, A., & Baillet, S. (2019). Brainstorm Pipeline Analysis of Resting-State Data From the Open MEG Archive. *Frontiers in Neuroscience*, *13*, 284. https://doi.org/10.3389/fnins.2019.00284

Norgaard, M., Matheson, G. J., Hansen, H. D., Thomas, A., Searle, G., Rizzo, G., Veronese, M., Giacomel, A., Yaqub, M., Tonietto, M., Funck, T., Gillman, A., Boniface, H., Routier, A., Dalenberg, J. R., Betthauser, T., Feingold, F., Markiewicz, C. J., Gorgolewski, K. J., … Ganz, M. (2021). PET-BIDS, an extension to the brain imaging data structure for positron emission tomography. *BioRxiv*, 2021.06.16.448390. https://doi.org/10.1101/2021.06.16.448390

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Obeid, I., & Picone, J. (2016). The Temple University Hospital EEG Data Corpus. *Frontiers in Neuroscience*, *10*, 196. https://doi.org/10.3389/fnins.2016.00196

Ojeda, A., Bigdely-Shamlo, N., & Makeig, S. (2014). MoBILAB: An open source toolbox for analysis and visualization of mobile brain/body imaging data. *Frontiers in Human Neuroscience*, *8*, 121. https://doi.org/10.3389/fnhum.2014.00121

Olvet, D. M., & Hajcak, G. (2009). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*(3), 156869. https://doi.org/10.1155/2011/156869

Palva, J. M., Wang, S. H., Palva, S., Zhigalov, A., Monto, S., Brookes, M. J., Schoffelen, J.-M., & Jerbi, K. (2018). Ghost interactions in MEG/EEG source space: A note of caution on inter-areal coupling measures. *NeuroImage*, *173*, 632–643. https://doi.org/10.1016/j.neuroimage.2018.02.032

Pardo, M. S., & Patterson, D. (2013). *Minds, Brains, and Law: The Conceptual Foundations*

of Law and Neuroscience. Oxford University Proess. https://doi.org/10.1093/acprof:oso/9780199812134.001.0001

Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, *21*(4), 474–483. https://doi.org/10.1038/s41593-018-0108-2

Paul, M., Govaart, G. H., & Schettino, A. (2021). Making ERP research more transparent: Guidelines for preregistration. *International Journal of Psychophysiology*, *164*, 52–63. https://doi.org/10.1016/j.ijpsycho.2021.02.016

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., Giorgio, G., He, X., Hinojosa, J. A., … Mushtaq, F. (2021). #EEGManyLabs: Investigating the Replicability of Influential EEG Experiments. *Cortex*. https://doi.org/10.1016/j.cortex.2021.03.013

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Pernet, C., 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., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2018). *Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG*. OSF Preprints. https://doi.org/10.31219/osf.io/a8dhx

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*, *23*(12), 1473–1483. https://doi.org/10.1038/s41593-020-00709-0

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. https://doi.org/10.1111/1469-8986.3720127

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*(6), 547–558.

https://doi.org/10.1111/j.1469-8986.1993.tb02081.x

Pizzo, F., Roehri, N., Medina Villalon, S., Trébuchon, A., Chen, S., Lagarde, S., Carron, R., Gavaret, M., Giusiano, B., McGonigal, A., Bartolomei, F., Badier, J. M., & Bénar, C. G. (2019). Deep brain activities can be detected with magnetoencephalography. *Nature Communications*, *10*(1), 971. https://doi.org/10.1038/s41467-019-08665-5

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

Poldrack, R. A., Whitaker, K., & Kennedy, D. (2020). Introduction to the special issue on reproducibility in neuroimaging. *NeuroImage*, *218*, 116357. https://doi.org/10.1016/j.neuroimage.2019.116357

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C.-T., Themanson, J. R., & Hillman, C. H. (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology*, *47*(4), 767–773.

Proctor, R. N. (1991). *Value-Free Science? Purity and Power in Modern Knowledge*. Harvard University Press.

Puce, A., & Hämäläinen, M. (2017). A Review of Issues Related to Data Acquisition and Analysis in EEG/MEG Studies. *Brain Sciences*, *7*(12), 58. https://doi.org/10.3390/brainsci7060058

Puoliväli, T., Palva, S., & Palva, J. M. (2020). Influence of multiple hypothesis testing on reproducibility in neuroimaging research: A simulation study and Python-based software. *Journal of Neuroscience Methods*, *337*, 108654. https://doi.org/10.1016/j.jneumeth.2020.108654

Quinn, A. J., Lopes-dos-Santos, V., Dupret, D., Nobre, A. C., & Woolrich, M. W. (2021). EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python. *Journal of Open Source Software*, *6*(59), 2977. https://doi.org/10.21105/joss.02977

Renkewitz, F., & Heene, M. (2019). The Replication Crisis and Open Science in Psychology. *Zeitschrift Für Psychologie*, *227*(4), 233–236. https://doi.org/10.1027/2151-2604/a000389

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., & Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, *3*(3), 199–217. https://doi.org/10.1038/s42256-021-00307-0

Rodu, J., & Kafadar, K. (2021). The q–q Boxplot. *Journal of Computational and Graphical Statistics*, *0*(0), 1–14. https://doi.org/10.1080/10618600.2021.1938586

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, *45*(6), 906–919. https://doi.org/10.1111/j.1469-8986.2008.00708.x

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering*, *16*(5), 051001. https://doi.org/10.1088/1741-2552/ab260c

Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The Quantitative Methods for Psychology*, *16*(4), 376–390. https://doi.org/10.20982/tqmp.16.4.p376

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., & Hadsell, R. (2016). Progressive Neural Networks. *ArXiv:1606.04671 [Cs]*. http://arxiv.org/abs/1606.04671

Salo, P., & Heikkinen, H. (2018). Slow Science: Research and Teaching for Sustainable Praxis. *Confero*, *6*(1). https://doi.org/10.3384/confero.2001-4562.181130

Sandre, A., Banica, I., Riesel, A., Flake, J., Klawohn, J., & Weinberg, A. (2020). Comparing the effects of different methodological decisions on the error-related negativity and its association with behaviour and gender. *International Journal of Psychophysiology*, *156*, 18–39. https://doi.org/10.1016/j.ijpsycho.2020.06.016

Santiuste, M., Nowak, R., Russi, A., Tarancon, T., Oliver, B., Ayats, E., Scheler, G., & Graetz, G. (2008). Simultaneous Magnetoencephalography and Intracranial EEG Registration: Technical and Clinical Aspects. *Journal of Clinical Neurophysiology*, *25*(6), 331–339. https://doi.org/10.1097/WNP.0b013e31818e7913

Sanz Leon, P., Knock, S. A., Woodman, M. M., Domide, L., Mersmann, J., McIntosh, A. R., & Jirsa, V. (2013). The Virtual Brain: A simulator of primate brain network dynamics. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00010

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211007468. https://doi.org/10.1177/25152459211007467

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

Schreiweis, C., Volle, E., Durr, A., Auffret, A., Delarasse, C., George, N., Dumont, M., Hassan, B. A., Renier, N., Rosso, C., Thiebaut de Schotten, M., Burguière, E., & Zujovic, V. (2019). A neuroscientific approach to increase gender equality. *Nature Human Behaviour*, *3*(12), 1238–1239. https://doi.org/10.1038/s41562-019-0755-7

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., Blankertz, B., & Haynes, J.-D. (2016). The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*, *113*(4), 1080–1085. https://doi.org/10.1073/pnas.1513569112

Seeber, M., Cantonas, L.-M., Hoevels, M., Sesia, T., Visser-Vandewalle, V., & Michel, C. M. (2019). Subcortical electrophysiological activity is detectable with high-density EEG source imaging. *Nature Communications*, *10*(1), 753. https://doi.org/10.1038/s41467-019-08725-w

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger Causality Analysis in Neuroscience and Neuroimaging. *Journal of Neuroscience*, *35*(8), 3293–3297. https://doi.org/10.1523/JNEUROSCI.4399-14.2015

Shafto, M. a, Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgleish, T., Henson, R. N., Brayne, C., & Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, 204. https://doi.org/10.1186/s12883-014-0204-1

Shaw, C., & Ward, L. (2014, March 6). *Dark thoughts: Why mental illness is on the rise in academia*. The Guardian. http://www.theguardian.com/higher-education-network/2014/mar/06/mental-health-academics-growing-problem-pressure-university

Simmons, A. M., & Luck, S. J. (2020). Protocol for Reducing COVID-19 Transmission Risk in EEG Research. *Research Square*, rs.3.pex-974. https://doi.org/10.21203/rs.3.pex-974/v2

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-Curve: A Key to the File-Drawer. *Journal of Experimental Psychology-General*, *143*(2), 534–547.

Sinha, S. R., Sullivan, L., Sabau, D., San-Juan, D., Dombrowski, K. E., Halford, J. J., Hani, A. J., Drislane, F. W., & Stecker, M. M. (2016). American Clinical Neurophysiology Society Guideline 1: Minimum Technical Requirements for Performing Clinical Electroencephalography. *Journal of Clinical Neurophysiology: Official Publication of*

the American Electroencephalographic Society, *33*(4), 303–307. https://doi.org/10.1097/WNP.0000000000000308

Sinnott-Armstrong, W. (2011). Lessons from Libet. In Conscious Will and Responsibility. In *Conscious Will and Responsibility*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195381641.001.0001/acprof-9780195381641-chapter-20

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384.

Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Kovic, V. (2021). How to do better N400 studies: Reproducibility, consistency and adherence to research standards in the existing literature. *Neuropsychology Review*. https://doi.org/10.1007/s11065-021-09513-4

Šoškić, A., Kappenman, E. S., Styles, S. J., & Ković, V. (2019). *Garden of forking paths in ERP research – effects of varying processing and analysis steps in an N400 experiment*. https://osf.io/6nqxy/

Šoškić, A., Ković, V., Ke, H., & Styles, S. J. (2020). *ARTEM-IS: Agreed Reporting Template for EEG Methodology - International Standard*. https://doi.org/10.17605/OSF.IO/PVRN6

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.

Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why Has the Number of Scientific Retractions Increased? *PLOS ONE*, *8*(7), e68397. https://doi.org/10.1371/journal.pone.0068397

Stengers, I. (2013). *Another Science is Possible. A Manifesto for Slow Science* (Muecke S. translation). Polity Press.

Stenner, T., Boulay, C., Grivich, M., Medine, D., Kothe, C., Grimm, G., xloem, tobiasherzke, Biancarelli, A., Mansencal, B., Frey, J., kyucrane, Powell, S., Clisson, P., & phfix. (2021). *sccn/liblsl: V1.15.1 - ASIO updated*. Zenodo. https://doi.org/10.5281/zenodo.5415959

Stewart, A. X., & Luck, S. J. (2020). *Standardized Measurement Error (SME) demo scripts*. https://osf.io/a4huc/

Styles, S. J., Kovic, V., Ke, H., & Šoškić, A. (2020). *ARTEM-IS Template for ERP - Draft for Consultation*. OSF. https://osf.io/w4nt6/

Styles, S. J., Kovic, V., Ke, H., & Šoškić, A. (2021). *Towards ARTEM-IS: An evidence-based Agreed Reporting Template for EEG Methodology - International Standard* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/myn7t

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), e2000797. https://doi.org/10.1371/journal.pbio.2000797

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

Tadel, F., Bock, E., Niso, G., Mosher, J. C., Cousineau, M., Pantazis, D., Leahy, R. M., & Baillet, S. (2019). MEG/EEG group analysis with Brainstorm. *Frontiers in Neuroscience*, *13*, 76.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition: High-pass filtering and artifactual ERP effects. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, X., & Henson, R. N. (2015). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, *144*, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Tenke, C. E., & Kayser, J. (2012). Generator localization by current source density (CSD): Implications of volume conduction and field closure at intracranial and scalp resolutions. *Clinical Neurophysiology*, *123*(12), 2328–2345. https://doi.org/10.1016/j.clinph.2012.06.005

Tibon, R., Open Science Committee, C., & Henson, R. (2018). Title TBA: Revising the Abstract Submission Process. *Trends in Cognitive Sciences*, *22*(4), 271–274. https://doi.org/10.1016/j.tics.2018.01.008

Turing, A. M. (1950). Computing Machinery and Intelligence. In *The Turing Test: Verbal Behavior as the Hallmark of Intelligence* (pp. 29–56). MIT Press.

van Vliet, M. (2019). Seven quick tips for data analysis pipelines in neuroimaging. *PLOS Computational Biology*, in press.

Vazire, S. (2017). Quality Uncertainty Erodes Trust in Science. *Collabra: Psychology*, *3*(1). https://doi.org/10.1525/collabra.74

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science*, *4*(3), 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Ware, C. (2019). *Information Visualization: Perception for Design*. Morgan Kaufmann.

Whiteford, K. L., Baltzell, L. S., Cooper, J. K., Irsik, V. C., Irvine, A., Mesik, J., Nolan, T., Oakes, B., Reed, A., Schrlau, A. E., Van Hedger, S., Bharadwaj, H., Johnsrude, I., Kidd Jr., G., Lim, S.-J., Luebke, A. E., Maddox, R. K., Marvin, E. W., Perrachione, T., … Oxenham, A. (2020). *Association of musical training with auditory and speech neural coding and perception*. https://doi.org/10.17605/OSF.IO/DUQ34

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*(Supplement C), 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., & Wilson, P. (2014). Best Practices for Scientific Computing. *PLOS Biology*, *12*(1), e1001745. https://doi.org/10.1371/journal.pbio.1001745

Wolpaw, J., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology*, *113*(6), 767–7. https://doi.org/10.1016/S1388-2457(02)00057-3

Wolpaw, J., & Wolpaw, E. W. (Eds.). (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press.

Xiong, C., Ceja, C. R., Ludwig, C. J. H., & Franconeri, S. (2020). Biased Average Position Estimates in Line and Bar Graphs: Underestimation, Overestimation, and Perceptual Pull. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 301–310. https://doi.org/10.1109/TVCG.2019.2934400

Yael, D., Vecht, J. J., & Bar-Gad, I. (2018). Filter-based phase shifts distort neuronal timing information. *Eneuro*, *5*(2).

Yarkoni, T., Markiewicz, C., de la Vega, A., Gorgolewski, K., Salo, T., Halchenko, Y., McNamara, Q., DeStasio, K., Poline, J.-B., Petrov, D., Hayot-Sasson, V., Nielson, D., Carlin, J., Kiar, G., Whitaker, K., DuPre, E., Wagner, A., Tirrell, L., Jas, M., … Blair, R. (2019). PyBIDS: Python tools for BIDS datasets. *Journal of Open Source Software*, *4*(40), 1294. https://doi.org/10.21105/joss.01294

Ylijoki, O.-H., & Mäntylä, H. (2003). Conflicting Time Perspectives in Academic Work. *Time & Society*, *12*(1), 55–78. https://doi.org/10.1177/0961463X03012001364

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient

Induced Gamma-Band Response in EEG as a Manifestation of Miniature Saccades. *Neuron*, *58*(3), 429–441. https://doi.org/10.1016/j.neuron.2008.03.027

Zander, T. O., & Kothe, C. A. (2011). Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, *8*(2), 025005. https://doi.org/10.1088/1741-2560/8/2/025005

Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences*, *113*(52), 14898–14903. https://doi.org/10.1073/pnas.1605155114

Zgraggen, E., Zhao, Z., Zeleznik, R., & Kraska, T. (2018). Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3174053