The replication crisis has led to positive structural, procedural, and community changes

Max Korbmacher[1,2,3], Flavio Azevedo[4*], Charlotte R. Pennington[5], Helena Hartmann[6], Madeleine Pownall[7], Kathleen Schmidt[8], Mahmoud Elsherif[9], Nate Breznau[10], Olly Robertson[11], Tamara Kalandadze[12], Shijun Yu[9], Bradley J. Baker[13], Aoife O'Mahony[14], Jørgen Ø.-S. Olsnes, John J. Shaw[15], Biljana Gjoneska[16], Yuki Yamada[17], Jan P. Röer[18], Jennifer Murphy[19], Shilaan Alzahawi[20], Sandra Grinschgl[21], Catia M. Oliveira[22], Tobias Wingen[23], Siu Kit Yeung[24], Meng Liu[25], Laura M. König[26], Nihan Albayrak-Aydemir[27,28], Oscar Lecuona[29,30], Leticia Micheli[31], Thomas Evans[32,33]

[1*]Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway.
[2]NORMENT Centre for Psychosis Research, University of Oslo and Oslo University Hospital, Oslo, Norway.
[3]Mohn Medical Imaging and Visualisation Center, Bergen, Norway.
[4]Department of Psychology, University of Cambridge, Cambridge, UK.
[5]School of Psychology, Aston University, Birmingham, UK.
[6]Department of Neurology, University of Essen, Essen, Germany.
[7]School of Psychology, University of Leeds, Leeds, UK.
[8]Department of Psychology, Ashland University, Ashland, USA.
[9]School of Psychology, University of Birmingham, Birmingham, UK.
[10]SOCIUM Research Center on Inequality and Social Policy, University of Bremen, Bremen, Germany.
[11]Department of Psychiatry, University of Oxford, Oxford, UK.
[12]Department of Education, ICT and Learning, Ostfold University College, Halden, Norway.
[13]Department of Sport and Recreation Management, Temple University, Philadelphia, USA.
[14]School of Psychology, Cardiff University, Cardiff, UK.
[15]Division of Psychology, De Montfort University, Leicester, UK.

[16]Macedonian Academy of Sciences and Arts, Skopje, North Macedonia.
[17]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan.
[18]Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany.
[19]Department of Applied Science, Technological University Dublin, Dublin, Ireland.
[20]Graduate School of Business, Stanford University, Standford, USA.
[21]nstitute of Psychology, University of Graz, Graz, Austria.
[22]Department of Psychology, University of York, York, UK.
[23]Institute of General Practice and Family Medicine, University of Bonn, Bonn, Germany.
[24]Department of Psychology, Chinese University of Hong Kong, Hong Kong, China.
[25]Faculty of Education, University of Cambridge, Cambridge, UK.
[26]Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Bayreuth, Germany.
[27]Open Psychology Research Centre, Open University, Walton Hall, UK.
[28]Department of Psychological and Behavioural Science, London School of Economics and Political Science, London, UK.
[29]Department of Psychology, Universidad Rey Juan Carlos, Madrid, Spain.
[30]Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain.
[31]Institute of Psychology, Leiden University, Leiden, The Netherlands.
[32]School of Human Sciences, University of Greenwich, Greenwich, UK.
[33]Institute for Lifecourse Development, University of Greenwich, Greenwich, UK.

*Corresponding author(s). E-mail(s): fa441@cam.ac.uk;

Abstract

In response to large-scale replication projects yielding successful replication rates substantially lower than expected, the behavioural, cognitive and social sciences have found themselves amidst a 'replication crisis'. In this narrative review, we reframe this 'crisis' through the lens of a credibility revolution, focusing on positive structural, procedural and community-driven changes. Second, we outline a path to expand ongoing advances and improvements. The credibility revolution has been an impetus to several substantive changes which will have a positive, long-term impact upon our research environment.

1 Introduction

"The history of science, like the history of all human ideas, is a history of irresponsible dreams, of obstinacy, and of error. But science is one of the very few human activities—perhaps the only one—in which errors are systematically criticised and fairly often, in time, corrected." - Karl Popper [1]

1.1 The replication crisis in psychology and the social sciences

After several notable scandals and controversies in 2011 [2–4], skepticism regarding claims in psychological science increased, and inspired the development of projects examining the replicability and reproducibility of past findings [5]. Replication refers to the process of repeating a study or experiment with the goal of verifying effects or generalising findings across new models or populations, whereas reproducibility refers to assessing the accuracy of the research claims based on the original methods, data, and/or code (see Table 1 for definitions).

- Insert Table 1 here -

In the most impactful replication initiative of the last decade, the Open Science Collaboration [6] sampled studies from three prominent journals representing different sub-fields of psychology to estimate the replicability of psychological research. Out of 100 independently performed replications, only 39% were subjectively labelled as successful replications, and on average, the effects were roughly half the original size. Putting these results into wider context, a minimum replicability rate of 89% should have been expected if all of the original effects were true (and not false positives; [7]). Pooling the Open Science Collaboration [6] replications with 207 other replications from recent years resulted in a higher estimate; 64% of effects successfully replicated with an effect size reduction of 32% [8]. Seemingly, estimations of replicability vary widely but nevertheless appear to be suboptimal. These suboptimal estimates of replicability are not exclusive to psychology, as they have been found across many other disciplines (e.g., animal behaviour; [9–11]; cancer biology [12]; economics, [13]), and appear to be symptomatic of persistent issues within the research environment [14, 15].

The 'replication crisis' has presented a number of considerable challenges to research, including concerns that it has undermined the role of science/scientists as a robust source of evidence for informing policy and practice [16], and has compromised the public's trust in science [17]. Simultaneously, the crisis has also presented a huge opportunity for development and reform. In this narrative review, we focus on the latter, exploring the replication crisis through the lens of a credibility revolution [18] to provide an overview of recent developments that have led to positive changes in the research landscape. We classify these into a) structural, b) procedural, and c) community change (see Fig.1), and discuss additional opportunities that the replication crisis is yet to fully bear.

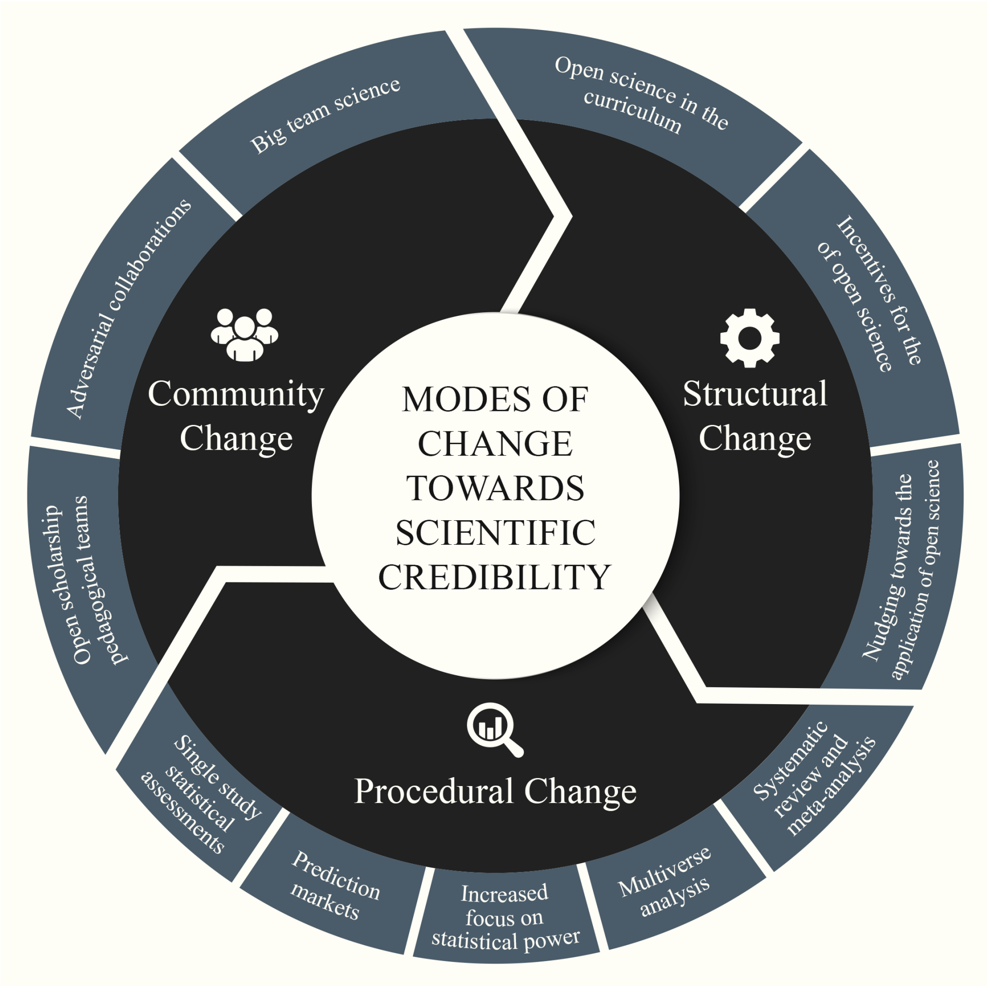

Fig. 1 Modes of change towards scientific credibility. This figure presents an overview of the three modes of change proposed in this article: structural change is often evoked at the institutional level expressed by new norms and rules; procedural change refers to single behaviours and sets of commonly used practices in the research process; community change encompasses how work and collaboration within the scientific community transforms.

1.2 Background: The replication crisis as a credibility revolution

Recent discussions have outlined various reasons why replications fail (see Box 1). To address these replicability concerns, different perspectives have been offered on how to reform, and promote improvements to, existing research norms in psychological science [19–21]. An academic movement collectively known as Open Scholarship (incorporating Open Science and Open Research) has driven positive change by accelerating the uptake of robust research practices, and advocating for a more diverse, equitable, inclusive, and accessible psychological science [22, 23].

These reforms have been driven by a diverse range of institutional initiatives, grass-roots, bottom-up initiatives and individuals. The extent of such impact led Vazire [18] to reframe the replicability crisis as a credibility revolution, acknowledging the term crisis reflects neither the intense self-examination of research disciplines in the last decade nor the various advances that have been implemented as a result.

- Insert Box 1 here -

Scientific practices are behaviours [24] and, in this way, they can be changed, especially when structures (e.g., funding agencies), environments (e.g., research groups), and individuals (e.g., single researchers) facilitate and support these changes. Most attempts to change the behaviours of individual researchers have concentrated on identifying or removing problematic practices and improving training in open scholarship [24]. As such, efforts to change individuals' behaviours have ranged from the creation of grass-roots communities to support individuals to incorporate open scholarship into their teaching (e.g., [25]), to infrastructural change (e.g., creation of open software (e.g., StatCheck [26] to identify statistical inconsistencies en-masse).

To tackle the substantive issues raised through the replication crisis requires much greater acknowledgement of the wider research landscape and culture, which includes the roles played by institutions, funders, and publishers [27]. For example, new 'open' journals have been developed, but they frequently operate in a marketplace which is based on reputation/prestige and thus uptake has been variable due to resistance from institutions and individuals. While there are various ongoing efforts from these stakeholder groups, they are often not integrated together, and their wider impact is infrequently evaluated. As such, whilst there have been many positive developments made at the structural and ground-level, there has been little consideration for how they have all contributed to a widespread change in the way science is considered, actioned and structured.

In this article we take the opportunity to reflect upon the scope and extent of positive changes as a result of the credibility revolution.To capture these different levels of change in our complex research landscape, we differentiate between a) structural, b) procedural, and c) community change. Our categorisation is not informed by any given theory, and there are overlaps and similarities across the outlined modes of change. However, this approach allows us to consider change in different domains: a) embedded norms, b) behaviours, and c) interactions, which we believe assists in demonstrating the scope of positive changes and which may allow us to empower and retain change-makers towards further positive scientific reform.

2 Structural Change

Structural change is often evoked at the institutional level, expressed by new norms and rules. In the following, we describe and discuss examples of structural change in the form of embedding open scholarship practices into the curriculum and by providing incentives.

2.1 Embedding replications into the curriculum

In the wake of the credibility revolution, higher education instructors and programs have begun to integrate open science practices into the curriculum at different levels. Most notably, some instructors have started including replications as part of basic research training and course curricula [28], and there are freely available, curated materials covering the entire process of executing replications with students [29] (see also forrt.org/reversals). In one prominent approach, the Collaborative Replications and Education Project [30, 31] integrates replications in undergraduate courses as coursework with a twofold goal: educating undergraduates to uphold high research standards whilst simultaneously advancing the field with replications. In this endeavour, the most cited studies from the most cited journals in the last three years serve as the sample from which students select their replication target. Administrative advisors then rate the feasibility of the replication to decide whether to run the study across the consortium of supervisors and students. After study completion, materials and data are submitted and used in meta-analyses, for which students are invited as co-authors.

In another proposed model [32], graduate students complete replication projects as part of their dissertation. Early career researchers (ECRs) are invited to prepare the manuscripts for publication [33–35] and, in this way, students' research efforts for their dissertation are utilised to contribute to a robust body of literature, while being formally acknowledged. An additional benefit is the opportunity for ECRs to further their career by publishing with available data. Institutions and departments can also benefit from embedding these projects as these not only increase the quality of education on research practices but also increase research outputs [22, 34, 36, 37].

If these models are to become commonplace, developing a set of standards regarding authorship would be beneficial. In particular, the question of what merits authorship can become an issue when student works are further developed, potentially without further involvement of the student. Such conflicts occur with other models of collaboration (see Community Change, below; [38]) but may be tackled by following standardized authorship templates, such as the Contributor Roles Taxonomy (CRediT) which helps detail each individual's contributions to the work [39, 40].

2.2 Wider embedding into curricula

In addition to embedding replications, open scholarship should be taught as a core component of higher education. Learning about open scholarship practices has been shown to influence student knowledge, expectations, attitudes, and engagement towards becoming more effective and responsible researchers, and consumers of science [41]. It is hence important to sufficiently address open scholarship in the classroom

[41], and different resources are available to help teaching staff with such implementations [29, 37]. Gaining an increased scientific literacy early on may have significant long-term benefits for students, including the opportunity to make a rigorous scientific contribution, acquiring a critical understanding of the scientific process and the value of replication, and a commitment to research values such as openness and transparency [41–43]. Embedding open research practices into education further shapes personal values that are connected to research, which will be crucial in later stages of both academic and non-academic careers [44]. This creates a path towards open scholarship values and practices becoming the norm, rather than the exception. It also links directly to existing movements often embraced among university students to foster greater equality and to break down status hierarchies which are fostered by gatekeepers in power positions (e.g., decolonisation efforts) [45].

Various efforts to increase the adoption of open scholarship practices into the curriculum are being undertaken by pedagogical teams with the overarching goal of increasing research rigor and transparency over time. While these changes are structural, they are often driven by single or small groups of individuals, often being in the early stages of their careers and receiving little recognition for their contributions [22]. An increasing number of grassroot open science organisations contribute in different educational roles, by providing resources and guidelines. The breadth of tasks required in the pedagogic reform towards open scholarship is exemplified by the Framework for Open and Reproducible Research Training (FORRT), focusing on reform and meta-scientific research to advance research transparency, reproducibility, rigour, social justice, and ethics [25]. The FORRT community is currently running more than 15 initiatives which include summaries of open scholarship literature, a crowdsourced glossary of open scholarship terms [46], a literature review of the impact on students of integrating open scholarship into teaching [41], out-of-the-box lesson plans [47], a team working on bridging neurodiversity and open scholarship [48, 49], and a living database of replications and reversals [50]. Other examples of organisations providing open scholarship materials are the Network of Open Science Initiatives [51] and Course Syllabi for Open and Reproducible Methods [52]. These and other collections of open-source teaching and learning materials (such as podcasts, how-to guides, courses, labs, open science networks, and databases) can facilitate the integration of open scholarship principles into education and practice. This not only raises awareness for open scholarship but also levels the playing field for researchers coming from countries or institutions with fewer resources, such as the 'Global South', referring to the regions outside of Europe and North America that are primarily politically and culturally marginalised such as regions in Asia and Africa [47, 53].

2.3 Incentives

Scientific practice is rooted in a deep history of problematic reward incentives (e.g., a focus on research quantity over quality, positive results) and so revised incentives are essential for a sustained integration of open scholarship practices. Current efforts have focused on the development of such incentives, which aim to target different actors, including students, academics, faculties and universities, funders, and journals [54–56]. As each of these actors have different and sometimes competing goals, the nature of the incentives to engage in open scholarship practices vary. In the following, we describe recently developed incentives targeting researchers, and academic journals and funders.

2.3.1 Targeting researchers

Traditional incentives for academics to achieve career opportunities and advancement are increased chances of publishing, winning grants, and signalling quality of the published work (e.g., perceived journal prestige) [45, 57]. In some journals, researchers are now given direct incentives for the preregistration of study plans and analyses before study execution, and openly accessible data and materials for their articles in the form of (open science) badges, with the aim of signalling study quality [24]. However, the extent to which badges can be used to increase open scholarship behaviours remains unclear; while one study [58] reports increased data sharing rates among articles published in Psychological Science, a recent randomized control trial no evidence for badges increasing data sharing [59], suggesting that more effective incentives or complementary workflows are required to motivate researchers to engage in open research practices [60].

Furthermore, there are incentives provided for different open scholarship practices, such as using the Registered Report publishing format [61, 62]. Here, authors submit research protocols for peer-review before data collection and/or analyses (in the case of secondary data). Registered Reports meeting high scientific standards are given provisional acceptance ('in-principle acceptance') prior to the results being known. Such format shifts the focus from the outcomes of the research to methodological quality, and accordingly realigns incentives by providing researchers with the certainty of publication, regardless of the nature of results, as long as they adhere to their preregistered protocol [61, 62]. Empirical evidence has also found that Registered Reports are perceived to be higher in research quality than regular articles, as well as equivalent in creativity and importance [63], which may provide further incentives for researchers to adopt this format.

2.3.2 Targeting journals and funders

Incentives are not limited to individual researchers but also to the general research infrastructure. One example of this is academic journals, which are attempting to implement standards to remain current and competitive, shown in the increase in open publishing options throughout the course of the credibility revolution and the newly formulated guidelines reinforcing these changes. For example, the Center for Open Science introduced the Transparency and Openness Promotion (TOP) Guidelines [64] which include eight modular standards to reflect journals' transparency standards; namely, citation standards, data transparency, analytic methods transparency, research materials transparency, design and analysis transparency, study preregistration, analysis plan preregistration, and replication. Building on these guidelines, the TOP factor quantifies the degree to which journals implement these standards, providing researchers with a guide on selecting journals accordingly. Similarly, organisations such as the UK Reproducibility Network [65], FORRT [66], NASA [67], UNESCO [68], and the European Commission [69] support open scholarship efforts publicly on an international level. Moreover, there are also efforts to open up funding options, for example, through the registered reports fundings scheme [36]. Here, funding allocation and publication review process are being combined into a single process, reducing the burden on reviewers and opportunities for questionable research practices. Finally, large-scale policies are being implemented supporting open scholarship practices, such as Plan-S. Plan-S mandates open publishing when the research is funded by public grants. The increase in open access options shows how journals are being effectively incentivized to expand their repertoire and normalize open access [70]. At the same time, article processing charges have increased, entailing the danger of excluding researchers from open access publishing who do not have the necessary funds to publish [70, 71]. As Plan-S shapes the decision space of journals and researchers, it is an incentive with the promise of long-term change.

Several efforts aim to re-design systems such as peer review and publishing. Community peer reviews (e.g., [72]) is a relatively new system in which experts review and recommend preprints to journals. Future developments in the direction of community peer review might contain an increased usage of overlay journals, meaning that the journals themselves do not manage their own content (including peer review) but rather select and curate content. The peer review procedures can also be changed, as the recent editorial decision in Elife to abolish accept/reject decisions during peer review has shown [73], and as reflected by a recommendation-based system of the community peer review system.

In addition, researchers in academic settings have largely been evaluated based on the number of papers they publish in high-impact journals [45], however see [57] for a criticism of journal impact factors) and their grant acquisition [36, 74]. In response, an increasing number of research stakeholders including universities have signed the San Francisco Declaration of Research Assessment (DORA); an initiative that aims to diversify the criteria used for hiring, promotion, and funding decisions by considering all research outputs, including software and data, as well as the qualitative impact of research (e.g., policy or practice implications). The goals of such initiatives are also reflected in efforts to foster sustainable change by regional hiring committees' requirements considering open scholarship practices [55], for an example see [75]). Such initiatives provide considerable incentives for individual researchers to follow open scholarship principles for personal career advancement.

3 Procedural Change

Procedural change refers to single behaviours and sets of commonly used practices in the research process. We describe and discuss prediction markets, statistical assessment tools, multiverse analysis, and systematic reviews and meta-analysis as examples of procedural changes.

3.1 Prediction markets of research credibility

In recent years, researchers have employed prediction markets to assess the credibility of research findings [76–81]. Here, researchers invite experts or non-experts to estimate the replicability of different studies or claims. Large prediction market projects such as the repliCATS project have yielded replicability predictions with high classification accuracy (between 61% and 86%, [78, 79]). The repliCATS project implemented a structured iterative evaluation procedure to solicit thousands of replication estimates which are now being used to develop prediction algorithms using machine learning. Though many prediction markets are composed of researchers or students with research training, even lay people seem to perform better than chance in predicting replicability [82]. Replication markets are positioned as complementary to, and an alternative to replication attempts, and there are various conditions where one approach may be preferable to implement (e.g., replication markets may be preferable where data collection is especially resource-intensive, but less so when study design may be especially complex). As such, the use of replication markets provides another tool available to researchers to provide insight on the credibility of existing and/or hypothetical works.

3.2 Statistical assessment tools

Failure to control error rates and design high-power studies can contribute to low replication rates [83, 84]. In response, researchers have developed various quantitative methods to assess expected distributions of statistical estimates (such as p-values), such as p-curving [85], or z-curving [86]. P-curve assesses publication bias by plotting the distribution of p-values across a set of studies, measuring the deviation from an expected uniform distribution of p-values considering a true null-hypothesis [85]. Similar to p-curve, z-curve assesses the distribution of test statistics while considering the power of statistical tests and false discovery rate within a body of literature [86]. Additionally, and perhaps most importantly, such estimations of bias in the literature identify selective reporting trends and help establish a better estimate of whether replication failures may be due to features of the original study or features of the replication study. Advocates of these methods argue for decreasing α-levels (i.e., the probability of finding a false positive / committing a type I error) when the likelihood of publication bias is high to allow for increased power and confidence in findings. Other researchers have called for reducing α-levels for all tests (e.g., from .05 to .005 [87]), rethinking null hypothesis statistical testing (NHST) and considering exploratory NHST [88], or abandoning NHST altogether ([89], see for an example [90]). However, these approaches are not panaceas and are unlikely to address all the highlighted concerns [20, 91]; instead, researchers have recommended simply justifying the alpha for tests with regard to the magnitude of acceptable Type I versus Type II (false negative) errors [91]. In this context, equivalence testing [92] or Bayesian analyses [93] have been proposed as suitable approaches to directly assess evidence for the alternative hypothesis against evidence for the null hypothesis [94]. GUI-based (point and click) statistical software packages such as JASP [95] or Jamovi [96], and the promotion of such methods including practical walkthroughs, shinyapps and Excel spreadsheets [92, 93] have made methods such as equivalence tests and Baysian statistics accessible to a wider audience of scientists.

3.3 Single study statistical assessments

A range of useful tools have been developed to pursue open values. For example, the accuracy of reported findings may be assessed by running simple, automated error-checks, such as StatCheck [26]. Validation studies [26] reported high sensitivity (larger than 83%), specificity (larger than 96%), and accuracy (larger than 92%) of the tool and that multiple correction procedures were not the reason for error in most cases. Other such innovations include the Granularity-Related Inconsistency of Means (GRIM) test [97], which aims to evaluate the consistency of mean values of integer data (e.g., from Likert-type scales), considering sample size and number of items, or Sample Parameter Reconstruction via Iterative TEchniques (SPRITE), which reconstructs samples and estimates of the item value distributions based on reported descriptive statistics [98]. Such techniques can serve as an initial step when reviewing the existing literature to ensure that a finding did not result from statistical errors or potential falsification, and can be implemented by the researchers themselves to check their work. With wider awareness and use of such tools, we have greater accessibility and capability to identify unsubstantiated claims.

3.4 Multiverse analysis

The multitude of researcher degrees of freedom (decisions researchers can make when using data) have been shown to influence the outcomes of analyses performed on the same data [99, 100]. In one investigation, 70 independent research teams analysed the same nine hypotheses with one neuroimaging dataset and results show data cleaning and statistical inferences varied considerably between teams; for example, no two groups used the same pipeline to pre-process the imaging data, which ultimately influenced both results and inference drawn from the results [99]. To increase transparency regarding researcher degrees of freedom, multiverse analyses can be performed [101, 102]. Conceptually, multiverse analyses consider the range of results obtained when all feasible and reasonable analyses are performed on the same dataset. While not all possible analyses conducted may be practically useful to include in a multiverse analysis, a range of appropriate analyses can provide evidence of a finding's robustness when pooled [103]. This approach allows researchers to simultaneously test the same hypothesis across a broad range of scenarios and determine the stability of certain effects as they navigate the often large 'garden of forking paths' [104].

3.5 Systematic review and meta-analysis

Systematic reviews or meta-analyses are used to synthesise findings from several primary studies [105, 106], which can reveal nuanced aspects of the research while keeping a bird's-eye perspective, for example, by presenting the range of effect sizes and resulting power estimates. Methods have been developed to assess the extent of publication bias in meta-analyses, and, to an extent, correct for it, using methods such as funnel plot asymmetry tests [107]. However, there are additional challenges influencing the results of meta-analyses and systematic reviews and hence their replicability, such as researcher degrees of freedom in determining inclusion criteria, methodological approaches, and the rigour of the primary studies [108, 109]. Thus, researchers have developed best practices for unbiased, open, and reproducible systematic reviews and meta-analyses such as Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [110], Non-Interventional, Reproducible, and Open (NIRO) [111] systematic review guidelines, the Generalized Systematic Review Registration Form [112], and PROSPERO, a register of systematic review protocols [113]. These guides and resources provide opportunities for more systematic accounts of research [114]. These guidelines often include a risk of bias assessment, where different biases are assessed and reported [115]. Yet, most systematic reviews and meta-analyses do not follow standardized reporting guidelines, even when required by the journal and stated in the article, reducing the reproducibility of primary and pooled effect size estimates [116]. An evaluated trial of enhanced requirements by some journals as part of the submission process [117] did lead to a slowly increasing uptake in such practices [118], with later findings indicating that protocol registrations increased the quality of associated meta-analyses [119]. Optimistically, continuous efforts to increase transparency appear to have already contributed to researchers more consistently reporting eligibility criteria, effect size information and synthesis techniques [120].

4 Community Change

Community change encompasses how work and collaboration within the scientific community transforms. We describe two of these developments in the following: big team science and adversarial collaborations.

4.1 Big team science

The credibility revolution has undoubtedly driven the formation and development of various large-scale, collaborative communities. Community examples of such approaches include mass replications, which as we discuss above can be integrated in research training [30, 31, 34, 121], and projects conducted by large teams and organisations such as the Many Labs studies [122, 123], the Hagen Cumulative Science Project [124], the Psychological Science Accelerator [125], and the Framework for Open and Reproducible Research Training (FORRT) [25]. A promising initiative to accelerate scientific progress is Big Team Science, i.e., large-scale collaborations of scientists working on a scholarly goal together, and pool resources across labs, institutions, disciplines, cultures, and countries [15, 126, 127]. Such groups have commonly formed to complete replication studies, with many disseminating their procedural learning in addition to scientific contributions [128]. Collaboration in large teams allows scientists to harness the expertise of a consortium of researchers, increase the efficiency of research resources (e.g., time and funding), and draw conclusions from richer cross-cultural samples [22, 126, 129]. Big team science typically involves a wide range of practices that enhance research quality, such as providing interdisciplinary internal reviews, multiple perspectives and viewpoints, implementing uniform and standardised protocols across participating laboratories, and recruiting larger and more diverse samples [41, 46, 125, 126, 128, 130]. The latter also extends to researchers themselves; big team science can increase representation, diversity and equality and allow researchers to collaborate by either coordinating data collection efforts at their respective institutions or by funding the data collection of researchers who may not have access to funds [131].

4.2 Adversarial collaborations

Scholarly critique typically occurs after research has been completed, for example, during peer-review or in back-and-forth commentaries of published work. With some exceptions [132–136], rarely do researchers who support contradictory theoretical frameworks work together to formulate research questions and design studies to test them. 'Adversarial collaborations' of this kind are arguably one of the most important developments in procedures to advance research because they allow for a consensus-based resolution of scientific debates and they facilitate more efficient knowledge production and self-correction by reducing bias [4, 137]. An example is the Transparent PSI Project [138] which united teams of researchers both supportive and critical of the idea of extra-sensory perception, allowing for a constructive dialogue and more agreeable consensus in conclusion.

A related practice to adversarial collaborations is that of 'red teams', which can be applied by both larger and smaller teams of researchers playing 'devil's advocate' between one another. Red teams work together to constructively criticise each other's work or to find errors during (but preferably early in) the entire research process, with the overarching goal of maximising research quality (Lakens, 2020). By avoiding errors "before it is too late", red teams have the potential to save large amounts of resources [126]. However, such initiatives are not without their issues. For example: A researcher has negative biases towards a specific marginalized group. The researcher proposes an adversarial collaboration with another researcher with similar biases but differing professional opinions or methodological expertise, knowing that both share the same biases. When publishing or promoting the work, the biases remain unchallenged. That way, adversarial collaborations can be used as a pretence to promote personal opinions including fringe scientific works under the veneer of a 'debate' (e.g., [139]).

5 Expanding structural, procedural and community changes

To expand the developments discussed, and to address current challenges in the field, we now highlight a selection of areas that can benefit from the previously described structural, procedural and community changes, namely: a) generalizability, b) theory building, and c) open scholarship for qualitative research, and end on d) diversity and inclusion as an area necessary to be considered in the context of open scholarship.

5.1 Generalizability

In extant work, the generalizability of effects is a serious concern (e.g., [140, 141]). Psychological researchers have traditionally focused on individual-level variability and failed to consider variables such as stimuli, tasks, or contexts over which they wish to generalise. While accounting for methodological variation can be achieved in part through statistical estimation (e.g., including random effects of stimuli in models) or acknowledging and discussing study limitations, unmeasured variables, stable contexts, and narrow samples still present substantive challenges to the generalizability of results [141].

Possible solutions lay in big team science and collaboration. Scientific communities such as the Psychological Science Accelerator (PSA) have aimed to test the generalizability of effects across cultures and beyond the Global North (i.e. the affluent and rich regions of the world, for example, North America, Europe, and Australia; [125]). However, big team science projects tend to be conducted voluntarily with very few resources in order to understand the diversity of a specific phenomenon (e.g. [142]). The large samples required to detect small effects may make it difficult for single researchers from specific countries to achieve adequate power for publication. Large global collaborations, such as the PSA, can therefore contribute to avoiding wasted resources by conducting large studies instead of many small-sample studies which are less impactful and more difficult to publish [125]. At the same time, large collaborations might offer a chance to counteract geographical inequalities in research outputs [143]. However, such projects also tend to recruit only the most accessible (typically student) populations from their countries, thereby potentially perpetuating issues of representation and diversity. Yet, increased efforts of international teams of scientists offer opportunities to provide both increased diversity in the research team and the research samples, posing opportunities to increase generalisability at various stages of the research process.

5.2 Formal theory building

Researchers have suggested that the replication crisis is, in fact, a “theory crisis” [144]; low rates of replicability may be explained in part by improper testing of theory or failures to identify auxiliary theoretical assumptions [84, 141]. The verbal formulation of psychological theories and hypotheses cannot always be directly tested with inferential statistics; thus generalisations are made which are not supported by the data [140, 141]. Yarkoni [141] has recommended moving away from broad, unspecific claims and theories towards specific quantitative tests that are interpreted with caution, and increased weighting of qualitative and descriptive research. Others have suggested formalising theories as computational models and engaging in theory testing rather than null hypothesis significance testing [144]. Indeed, many researchers may not even be at a stage where they are ready or able to to test hypotheses [145]. Additional discussion of how to improve psychological theory and its evaluation is needed to advance the credibility revolution. Such discussions reassessing the application of statistics (in the context of statistical theory) are important steps in improving research quality.

5.3 Qualitative research

Open scholarship research and discussion has focused primarily on quantitative data collection and analyses, with substantively less consideration for compatibility with qualitative or mixed methods [47, 146]. Qualitative research presents methodological, ontological, epistemological, and ethical challenges that need to be considered to increase openness while preserving the integrity of the research process. The uniqueness, context-dependent, and labour-intensive features of qualitative research can create barriers, for example to preregistration or data sharing [147]. Similarly, some of the tools, practices, and concerns of open scholarship are simply not compatible with many qualitative epistemological approaches (e.g., a concern for replicability; [41]). Thus, a one-size-fits-all approach to qualitative or mixed methods data sharing and engagement with other open scholarship tools may not be appropriate for safeguarding the fundamental principles of qualitative research (see review [148]). However, there is a growing body of literature offering descriptions on how to engage in open scholarship practices when executing qualitative studies to move the field forward [7, 147, 149–151], and protocols developed specifically for qualitative research, such as preregistration templates [152]. Better representation of the application of open scholarship practices like a buffet, which can be chosen from, depending on the projects and its limitations and opportunities [7, 153] is ongoing. Such an approach is reflected in various studies describing the tailored application of open scholarship protocols in qualitative studies [149, 151]. Overall, validity, transparency, ethics, reflexivity, and collaboration can be fostered by engaging in qualitative open science; open practices which allow others to understand the research process and its knowledge generation are particularly impactful here [147, 150]. Irrespective of the methodological and epistemological approach, then, transparency is key to the effective communication and evaluation of results from both quantitative and qualitative studies and there have been promising developments within qualitative and mixed research towards increasing uptake of open scholarship practices.

5.4 Diversity and inclusion

An important point to consider when encouraging change is that the playing field is not equal for all actors, underlining the need for flexibility that takes into account regional differences and marginalised groups as well as differences in resource allocation when implementing open science practices [47]. For example, there are clear differences in the availability of resources by geographic region [154, 155] and social groups, by ethnicity [156] or sex and gender [47, 157, 158]. Resource disparities are also self-sustaining as, for instance, funding increases the chances of conducting research at or beyond the state-of-the-art which in turn increases the chances of obtaining future funding [154]. Choosing (preferably free) open access options, including preprints and postprints is one step to allow scholars to access resources irrespective of their priveleges. Other options are pooled funding applications, re-distributions of resources in international teams of researchers, and international collaborations. Big team science is a promising avenue to produce high-quality research while inviting diversity [126]; yet, the predominantly volunteering-based system of such team science might exclude researchers who do not have allocated hours or funding for such team efforts. Hence, beyond these procedural and community changes, structural change is also essential.

6 Outlook: what can we learn in the future?

Evidenced by the scale of developments discussed, the replication crisis has motivated structural, procedural and community changes that would have previously been considered idealistic and delayed through resistance. Importantly, while the developments within the credibility revolution were originally fuelled by failed replications, these in themselves are not the only threat to scientific credibility. For different fields, replication rates should be expected to vary and may not be an ideal measure of research credibility or quality, as replicability will depend on the questions the research addresses, experimental setups, and the objects of examination [159]. Thus, we should look for opportunities to maximise transparency, scientific rigour and quality wherever they exist [19]. As such, we encourage researchers to 'lean in' to the possibilities made available from 'crises' and their revolutions.

Learning from the credibility revolution, we recommend assessing crises holistically. Instead of focussing on single issues, structural, procedural and community processes should be observed as intertwined drivers of the credibility revolution. After identifying problems, actionable changes need to be implemented on all of these levels, requiring the actors in these different domains to actively take responsibility for improvements of the field [160]. If one is fixed without the other (e.g., researchers focus on high-quality outputs [*individual level*] but are incentivised to focus on novelty [*structural level*]), then the problems will prevail and reform will fail. From the presented changes, however, it becomes clear that there are multiple positive changes already implemented, offering additional opportunities for further advances to shape the way forward in the credibility revolution.

Acknowledgments. We want to thank Ali H. Al-Hoorie for his contribution with reference formatting and commenting the manuscript, and Sriraj Aiyer for the comments provided on an earlier version of the manuscript.

Declarations

- Conflict of interest/Competing interests: All authors are advocates of open scholarship and are members of various initiatives including the Framework for Open and Reproducible Research Training (FORRT), the Student Initiative for Open Science, ReplicabiliTEA journal clubs, and the United Kingdom Reproducibility Network (UKRN). ... MORE WILL BE ADDED DEPENDENT ON THE AUTHORSHIP SURVEY
- Authors contributions statement: Author contributions and roles are laid out based on the CRediT (Contributor Roles Taxonomy). Please visit https://www.casrai.org/credit.html for details and definitions of each of the roles listed below [161].

  Conceptualization: Max Korbmacher.
  Methodology: Max Korbmacher, Flavio Azevedo, and Thomas R. Evans.
  Project administration: Max Korbmacher and Flavio Azevedo.
  Visualisation: Max Korbmacher and Flavio Azevedo.

Writing - original draft: Max Korbmacher, Flavio Azevedo, Thomas R. Evans, Charlotte R. Pennington.
Writing - review and editing: Max Korbmacher, Flavio Azevedo, Helena Hartmann, Mahmoud M. Elsherif, Aoife O'Mahony, Tamara Kalandadze, Jørgen Ø.-S. Olsnes, Charlotte R. Pennington, John J. Shaw, Madeleine Pownall, Biljana Gjoneska, Yuki Yamada, Jan P. Röer, Jennifer Murphy, Shilaan Alzahawi, Bradley J. Baker, Sandra Grinschgl, Catia M. Oliveira, Tobias Wingen, Siu Kit Yeung, Meng Liu, Laura M. König, Nihan Albayrak-Aydemir, Oscar Lecuona, Thomas R. Evans, Olly Robertson, Leticia Micheli, Kathleen Schmidt, and Shijun Yu.

References

[1] Popper, K.R.: Conjectures and Refutations. The Growth of Scientific Knowledge.(Essays and Lectures.). Routledge & Kegan Paul, ??? (1963)

[2] Bem, D.J.: Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. Journal of Personality and Social Psychology **100**(3), 407 (2011)

[3] Crocker, J.: The road to fraud starts with a single step. Nature **479**(7372), 151–151 (2011)

[4] Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Van Der Maas, H.L.: Why psychologists must change the way they analyze their data: the case of psi: comment on bem (2011). (2011)

[5] Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., Ioannidis, J.: A manifesto for reproducible science. Nature human behaviour **1**(1), 1–9 (2017)

[6] OpenScienceCollaboration: Estimating the reproducibility of psychological science. Science **349**(6251), 4716 (2015)

[7] Field, S.M., Hoekstra, R., Bringmann, L., Ravenzwaaij, D.: When and why to replicate: As easy as 1, 2, 3? Collabra: Psychology **5**(1) (2019)

[8] Nosek, B.A., Hardwicke, T.E., Moshontz, H., Allard, A., Corker, K.S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M.B., *et al.*: Replicability, robustness, and reproducibility in psychological science. Annual review of psychology **73**, 719–748 (2022)

[9] Farrar, B.G., Boeckle, M., Clayton, N.S.: Replications in comparative cognition: what should we expect and how can we improve? Animal behavior and cognition **7**(1), 1 (2020)

[10] Farrar, B.G., Voudouris, K., Clayton, N.S.: Replications, comparisons, sampling

and the problem of representativeness in animal cognition research. Animal behavior and cognition **8**(2), 273 (2021)

[11] Farrar, B.G., Vernouillet, A., Garcia-Pelegrin, E., Legg, E.W., Brecht, K.F., Lambert, P.J., Elsherif, M., Francis, S., O'Neill, L., Clayton, N.S., *et al.*: Reporting and interpreting non-significant results in animal cognition research. PeerJ **11**, 14963 (2023)

[12] Errington, T.M., Mathur, M., Soderberg, C.K., Denis, A., Perfito, N., Iorns, E., Nosek, B.A.: Investigating the replicability of preclinical cancer biology. Elife **10**, 71601 (2021)

[13] Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., *et al.*: Evaluating replicability of laboratory experiments in economics. Science **351**(6280), 1433–1436 (2016)

[14] Frith, U.: Fast lane to slow science. Trends in cognitive sciences **24**(1), 1–2 (2020)

[15] Pennington, C.: A student's guide to open science: Using the replication crisis to reform psychology (2023)

[16] Sanders, M., Snijders, V., Hallsworth, M.: Behavioural science and policy: where are we now and where are we going? Behavioural Public Policy **2**(2), 144–167 (2018)

[17] Hendriks, F., Kienhues, D., Bromme, R.: Replication crisis= trust crisis? the effect of successful vs failed replications on laypeople's trust in researchers and research. Public Understanding of Science **29**(3), 270–288 (2020)

[18] Vazire, S.: Implications of the credibility revolution for productivity, creativity, and progress. Perspectives on Psychological Science **13**(4), 411–417 (2018)

[19] Freese, J., Rauf, T., Voelkel, J.G.: Advances in transparency and reproducibility in the social sciences. Social Science Research **107**, 102770 (2022)

[20] Trafimow, D., Amrhein, V., Areshenkoff, C.N., Barrera-Causil, C.J., Beh, E.J., Bilgiç, Y.K., Bono, R., Bradley, M.T., Briggs, W.M., Cepeda-Freyre, H.A., *et al.*: Manipulating the alpha level cannot cure significance testing. Frontiers in psychology **9**, 699 (2018)

[21] Loken, E., Gelman, A.: Measurement error and the replication crisis. Science **355**(6325), 584–585 (2017)

[22] Azevedo, F., Liu, M., Pennington, C.R., Pownall, M., Evans, T.R., Parsons, S., Elsherif, M.M., Micheli, L., Westwood, S.J., Open, R.R.T.F.f.o.: Towards a culture of open scholarship: the role of pedagogical communities. BMC Research

Notes **15**(1), 75 (2022)

[23] Grahe, J.E., Cuccolo, K., Leighton, D.C., Cramblet Alvarez, L.D.: Open science promotes diverse, just, and sustainable research and educational outcomes. Psychology Learning & Teaching **19**(1), 5–20 (2020)

[24] Norris, E., O'Connor, D.B.: Science as behaviour: Using a behaviour change approach to increase uptake of open science. Taylor & Francis (2019)

[25] Azevedo, F., Parsons, S., Micheli, L., Strand, J., Rinke, E.M., Guay, S., Elsherif, M., Quinn, K.A., Wagge, J., Steltenpohl, C.N., et al.: Introducing a framework for open and reproducible research training (forrt) (2019)

[26] Nuijten, M.B., Polanin, J.R.: "statcheck": Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. Research synthesis methods **11**(5), 574–579 (2020)

[27] Evans, T.R., Pownall, M., Collins, E., Henderson, E.L., Pickering, J.S., O'Mahony, A., Zaneva, M., Jaquiery, M., Dumbalska, T.: A network of change: united action on research integrity. BMC research notes **15**(1), 1–7 (2022)

[28] Koole, S.L., Lakens, D.: Rewarding replications: A sure and simple way to improve psychological science. Perspectives on Psychological Science **7**(6), 608–614 (2012)

[29] Bauer, G., Breznau, N., Gereke, J., Höffler, J.H., Janz, N., Rahal, R.-M., Rennstich, J.K., Soiné, H.: Teaching constructive replications in the social sciences (2022)

[30] Wagge, J.R., Baciu, C., Banas, K., Nadler, J.T., Schwarz, S., Weisberg, Y., IJzerman, H., Legate, N., Grahe, J.: A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect. Collabra: Psychology **5**(1) (2019)

[31] Wagge, J.R., Brandt, M.J., Lazarevic, L.B., Legate, N., Christopherson, C., Wiggins, B., Grahe, J.E.: Publishing research with undergraduate students via replication work: The collaborative replications and education project. Frontiers in psychology, 247 (2019)

[32] Quintana, D.S.: Replication studies for undergraduate theses to improve science and education. Nature Human Behaviour **5**(9), 1117–1118 (2021)

[33] Efendić, E., Chandrashekar, S.P., Lee, C.S., Yeung, L.Y., Kim, M.J., Lee, C.Y., Feldman, G.: Risky therefore not beneficial: Replication and extension of finucane et al.'s (2000) affect heuristic experiment. Social Psychological and Personality Science **13**(7), 1173–1184 (2022)

[34] Feldman, G.: Replications and extensions of classic findings in Judgment and Decision Making. Project overview at https://doi.org/10.17605/OSF.IO/5Z4A8 (2020)

[35] Ziano, I., Yao, J.D., Gao, Y., Feldman, G.: Impact of ownership on liking and value: Replications and extensions of three ownership effect experiments. Journal of Experimental Social Psychology **89**, 103972 (2020)

[36] Munafò, M.R.: Improving the efficiency of grant and journal peer review: registered reports funding. Nicotine & Tobacco Research **19**(7), 773–773 (2017)

[37] Pownall, M., Azevedo, F., Aldoh, A., Elsherif, M., Vasilev, M., Pennington, C.R., Robertson, O., Tromp, M.V., Liu, M., Makel, M.C., et al.: Embedding open and reproducible science into teaching: A bank of lesson plans and resources. Scholarship of Teaching and Learning in Psychology (2021)

[38] Coles, N.A., DeBruine, L.M., Azevedo, F., Baumgartner, H.A., Frank, M.C.: 'big team'science challenges us to reconsider authorship. Nature Human Behaviour, 1–3 (2023)

[39] Allen, L., O'Connell, A., Kiermer, V.: How can we ensure visibility and diversity in research contributions? how the contributor role taxonomy (credit) is helping the shift from authorship to contributorship. Learned Publishing **32**(1), 71–74 (2019)

[40] Allen, L., Scott, J., Brand, A., Hlava, M., Altman, M.: Publishing: Credit where credit is due. Nature **508**(7496), 312–313 (2014)

[41] Pownall, M., Azevedo, F., König, L.M., Slack, H.R., Evans, T., Flack, Z., Grinschgl, S., Elsherif, M.M., Gilligan-Lee, K.A., Oliveira, C.M.F., et al.: The impact of open and reproducible scholarship on students' scientific literacy, engagement, and attitudes towards science: A review and synthesis of the evidence (2022)

[42] Chopik, W.J., Bremner, R.H., Defever, A.M., Keller, V.N.: How (and whether) to teach undergraduates about the replication crisis in psychological science. Teaching of psychology **45**(2), 158–163 (2018)

[43] Frank, M.C., Saxe, R.: Teaching replication. Perspectives on Psychological Science **7**(6), 600–604 (2012)

[44] Levin, N., Leonelli, S.: How does one "open" science? questions of value in biological research. Science, Technology, & Human Values **42**(2), 280–305 (2017)

[45] Van Dijk, D., Manor, O., Carey, L.B.: Publication metrics and success on the academic job market. Current Biology **24**(11), 516–517 (2014)

[46] Parsons, S., Azevedo, F., Elsherif, M.M., Guay, S., Shahim, O.N., Govaart, G.H.,

Norris, E., O'mahony, A., Parker, A.J., Todorovic, A., *et al.*: A community-sourced glossary of open scholarship terms. Nature Human Behaviour **6**(3), 312–318 (2022)

[47] Pownall, M., Talbot, C.V., Henschel, A., Lautarescu, A., Lloyd, K.E., Hartmann, H., Darda, K.M., Tang, K.T., Carmichael-Murphy, P., Siegel, J.A.: Navigating open science as early career feminist researchers. Psychology of Women Quarterly **45**(4), 526–539 (2021)

[48] Elsherif, M.M., Middleton, S.L., Phan, J.M., Azevedo, F., Iley, B.J., Grose-Hodge, M., Tyler, S.L., Kapp, S.K., Gourdon-Kanhukamwe, A., Grafton-Clarke, D., et al.: Bridging neurodiversity and open scholarship: How shared values can guide best practices for research integrity, social justice, and principled education. MetaArXiv. June **20** (2022)

[49] Gourdon-Kanhukamwe, A., Kalandadze, T., Yeung, S.K., Azevedo, F., Iley, B., Phan, J.M., Ramji, A.V., Shaw, J.J., Zaneva, M., Dokovova, M., et al.: Opening up understanding of neurodiversity: A call for applying participatory and open scholarship practices (2022)

[50] Leech, G.: Reversals in psychology. Behavioural and Social Sciences at Nature Portfolio article at https://socialsciences.nature.com/posts/reversals-in-psychology (2021)

[51] https://osf.io/tbkzh/

[52] https://osf.io/vkhbt/

[53] Collyer, F.M.: Global patterns in the publishing of academic knowledge: Global north, global south. Current Sociology **66**(1), 56–73 (2018)

[54] Ali-Khan, S.E., Harris, L.W., Gold, E.R.: Motivating participation in open science by examining researcher incentives. Elife **6**, 29319 (2017)

[55] Robson, S.G., Baum, M.A., Beaudry, J.L., Beitner, J., Brohmer, H., Chin, J.M., Jasko, K., Kouros, C.D., Laukkonen, R.E., Moreau, D., *et al.*: Promoting open science: a holistic approach to changing behaviour. Collabra: Psychology **7**(1), 30137 (2021)

[56] https://coara.eu/all-resources/

[57] Vanclay, J.K.: Impact factor: outdated artefact or stepping-stone to journal certification? Scientometrics **92**(2), 211–238 (2012)

[58] Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., *et al.*: Badges to acknowledge open practices: A simple, low-cost, effective method

for increasing transparency. PLoS biology **14**(5), 1002456 (2016)

[59] Rowhani-Farid, A., Aldcroft, A., Barnett, A.G.: Did awarding badges increase data sharing in bmj open? a randomized controlled trial. Royal Society open science **7**(3), 191818 (2020)

[60] Thibault, R.T., Pennington, C.R., Munafo, M.: Reflections on preregistration: Core criteria, badges, complementary workflows (2022)

[61] Chambers, C.D.: Registered reports: A new publishing initiative at cortex. Cortex **49**(3), 609–610 (2013)

[62] Chambers, C.D., Tzavella, L.: The past, present and future of registered reports. Nature human behaviour **6**(1), 29–42 (2022)

[63] Soderberg, C.K., Errington, T.M., Schiavone, S.R., Bottesini, J., Thorn, F.S., Vazire, S., Esterling, K.M., Nosek, B.A.: Initial evidence of research quality of registered reports compared with the standard publishing model. Nature Human Behaviour **5**(8), 990–997 (2021)

[64] Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S., Breckler, S., Buck, S., Chambers, C., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D., Hesse, B., Humphreys, M., DeHaven, A.C.: Transparency and Openness Promotion (TOP) Guidelines. Center for Open Science project at https://osf.io/9f6gx/ (2022)

[65] https://www.ukrn.org/

[66] https://www.forrt.org/

[67] https://github.com/nasa/transform-to-open-science

[68] https://www.unesco.org/en/natural-sciences/open-science

[69] https://eua.eu/downloads/publications/eua%20os%20agenda.pdf

[70] Zhang, L., Wei, Y., Huang, Y., Sivertsen, G.: Should open access lead to closed research? the trends towards paying to perform research. Scientometrics **127**(12), 7653–7679 (2022)

[71] McNutt, M.: "plan s" falls short for society publishers—and for the researchers they serve. Proceedings of the National Academy of Sciences **116**(7), 2400–2403 (2019)

[72] https://peercommunityin.org/

[73] https://elifesciences.org/for-the-press/b2329859/elife-ends-accept-reject-decisions-following-peer-review

[74] Nosek, B.A., Spies, J.R., Motyl, M.: Scientific utopia: Ii. restructuring incentives and practices to promote truth over publishability. Perspectives on Psychological Science **7**(6), 615–631 (2012)

[75] https://www.nicebread.de/open-science-hiring-practices/

[76] Delios, A., Clemente, E.G., Wu, T., Tan, H., Wang, Y., Gordon, M., Viganola, D., Chen, Z., Dreber, A., Johannesson, M., *et al.*: Examining the generalizability of research findings from archival data. Proceedings of the National Academy of Sciences **119**(30), 2120377119 (2022)

[77] Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B.A., Johannesson, M.: Using prediction markets to estimate the reproducibility of scientific research. Proceedings of the National Academy of Sciences **112**(50), 15343–15347 (2015)

[78] Fraser, H., Bush, M., Wintle, B.C., Mody, F., Smith, E.T., Hanea, A.M., Gould, E., Hemming, V., Hamilton, D.G., Rumpff, L., *et al.*: Predicting reliability through structured expert elicitation with the replicats (collaborative assessments for trustworthy science) process. Plos one **18**(1), 0274429 (2023)

[79] Gordon, M., Viganola, D., Dreber, A., Johannesson, M., Pfeiffer, T.: Predicting replicability—analysis of survey and prediction market data from large-scale forecasting projects. Plos one **16**(4), 0248780 (2021)

[80] Tierney, W., Hardy III, J.H., Ebersole, C.R., Leavitt, K., Viganola, D., Clemente, E.G., Gordon, M., Dreber, A., Johannesson, M., Pfeiffer, T., *et al.*: Creative destruction in science. Organizational Behavior and Human Decision Processes **161**, 291–309 (2020)

[81] Tierney, W., Hardy III, J., Ebersole, C.R., Viganola, D., Clemente, E.G., Gordon, M., Hoogeveen, S., Haaf, J., Dreber, A., Johannesson, M., *et al.*: A creative destruction approach to replication: Implicit work and sex morality across cultures. Journal of Experimental Social Psychology **93**, 104060 (2021)

[82] Hoogeveen, S., Sarafoglou, A., Wagenmakers, E.-J.: Laypeople can predict which social-science studies will be replicated successfully. Advances in Methods and Practices in Psychological Science **3**(3), 267–285 (2020)

[83] Button, K.S., Munafò, M.R.: Powering reproducible research. Psychological science under scrutiny: Recent challenges and proposed solutions, 22–33 (2017)

[84] Światkowski, W., Dompnier, B.: Replicability crisis in social psychology: Looking at the past to find new pathways for the future. International Review of Social Psychology **30**(1), 111–124 (2017)

[85] Simonsohn, U., Nelson, L.D., Simmons, J.P.: P-curve: a key to the file-drawer.

Journal of experimental psychology: General **143**(2), 534 (2014)

[86] Brunner, J., Schimmack, U.: Estimating population mean power under conditions of heterogeneity and selection for significance. Meta-Psychology **4** (2020)

[87] Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C., *et al.*: Redefine statistical significance. Nature human behaviour **2**(1), 6–10 (2018)

[88] Rubin, M., Donkin, C.: Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. Philosophical Psychology, 1–29 (2022)

[89] Amrhein, V., Greenland, S.: Remove, rather than redefine, statistical significance. Nature human behaviour **2**(1), 4–4 (2018)

[90] Trafimow, D., Marks, M.: Editorial in basic and applied social pschology. Basic and Applied Social Pschology **37**, 1–2 (2015)

[91] Lakens, D., Adolfi, F.G., Albers, C.J., Anvari, F., Apps, M.A., Argamon, S.E., Baguley, T., Becker, R.B., Benning, S.D., Bradford, D.E., *et al.*: Justify your alpha. Nature human behaviour **2**(3), 168–171 (2018)

[92] Lakens, D., Scheel, A.M., Isager, P.M.: Equivalence testing for psychological research: A tutorial. Advances in Methods and Practices in Psychological Science **1**(2), 259–269 (2018)

[93] Verhagen, J., Wagenmakers, E.-J.: Bayesian tests to quantify the result of a replication attempt. Journal of Experimental Psychology: General **143**(4), 1457 (2014)

[94] Dienes, Z.: Bayesian versus orthodox statistics: Which side are you on? Perspectives on Psychological Science **6**(3), 274–290 (2011)

[95] Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q.F., Šmíra, M., Epskamp, S., *et al.*: Jasp: Graphical statistical software for common statistical designs. Journal of Statistical Software **88**, 1–17 (2019)

[96] Şahin, M., Aybek, E.: Jamovi: an easy to use statistical software for the social scientists. International Journal of Assessment Tools in Education **6**(4), 670–692 (2019)

[97] Brown, N.J., Heathers, J.A.: The grim test: A simple technique detects numerous anomalies in the reporting of results in psychology. Social Psychological and Personality Science **8**(4), 363–369 (2017)

[98] Heathers, J.A., Anaya, J., Zee, T., Brown, N.J.: Recovering data from summary

statistics: Sample parameter reconstruction via iterative techniques (sprite). Technical report, PeerJ Preprints (2018)

[99] Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., *et al.*: Variability in the analysis of a single neuroimaging dataset by many teams. Nature **582**(7810), 84–88 (2020)

[100] Breznau, N., Rinke, E.M., Wuttke, A., Nguyen, H.H., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H.K., Auer, D., Azevedo, F., *et al.*: Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proceedings of the National Academy of Sciences **119**(44), 2203150119 (2022)

[101] Azevedo, F., Jost, J.T.: The ideological basis of antiscientific attitudes: Effects of authoritarianism, conservatism, religiosity, social dominance, and system justification. Group Processes & Intergroup Relations **24**(4), 518–549 (2021)

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

[103] Wagenmakers, E.-J., Sarafoglou, A., Aczel, B.: One statistical analysis must not rule them all. Nature **605**(7910), 423–425 (2022)

[104] Gellman, A., Lokem, E.: The statistical crisis in science data-dependent analysis—a 'garden of forking paths'—explains why many statistically significant comparisons don't hold up. Am. Sci **102**, 460 (2014)

[105] Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R.: Introduction to Meta-analysis. John Wiley & Sons, ??? (2021)

[106] Higgins, J.P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A.: Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, ??? (2019)

[107] Carter, E.C., Schönbrodt, F.D., Gervais, W.M., Hilgard, J.: Correcting for bias in psychology: A comparison of meta-analytic methods. Advances in Methods and Practices in Psychological Science **2**(2), 115–144 (2019)

[108] Nuijten, M.B., Hartgerink, C.H., Van Assen, M.A., Epskamp, S., Wicherts, J.M.: The prevalence of statistical reporting errors in psychology (1985–2013). Behavior research methods **48**, 1205–1226 (2016)

[109] Van Assen, M.A., Aert, R., Wicherts, J.M.: Meta-analysis using effect size distributions of only statistically significant studies. Psychological methods **20**(3), 293 (2015)

[110] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., *et al.*: The prisma 2020 statement: an updated guideline for reporting systematic reviews. International journal of surgery **88**, 105906 (2021)

[111] Topor, M.K., Pickering, J.S., Mendes, A.B., Bishop, D., Büttner, F., Elsherif, M.M., Evans, T.R., Henderson, E.L., Kalandadze, T., Nitschke, F.T., et al.: An integrative framework for planning and conducting non-intervention, reproducible, and open systematic reviews (niro-sr). Meta-Psychology (2022)

[112] Akker, O., Peters, G.Y., Bakker, C., Carlsson, R., Coles, N.A., Corker, K.S., Feldman, G., Moreau, D., Nordstrøm, T., Pickering, J.S., Riegelman, A., Topor, M., Veggel, N., Yeung, S.K., Mellor, D., Pfeiffer, N.: Generalized systematic review registration form. OSF Preprint (2020)

[113] Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., Stewart, L.: The nuts and bolts of prospero: an international prospective register of systematic reviews. Systematic reviews **1**(1), 1–9 (2012)

[114] Cristea, I.A., Naudet, F., Caquelin, L.: Meta-research studies should improve and evaluate their own data sharing practices. Journal of Clinical Epidemiology **149**, 183–189 (2022)

[115] Knobloch, K., Yoon, U., Vogt, P.M.: Preferred reporting items for systematic reviews and meta-analyses (prisma) statement and publication bias. Journal of Cranio-Maxillofacial Surgery **39**(2), 91–92 (2011)

[116] Lakens, D., Hilgard, J., Staaks, J.: On the reproducibility of meta-analyses: Six practical recommendations. BMC psychology **4**, 1–10 (2016)

[117] Editors, P.M.: Best practice in systematic reviews: the importance of protocols and registration. PLoS medicine **8**(2), 1001009 (2011)

[118] Tsujimoto, Y., Tsujimoto, H., Kataoka, Y., Kimachi, M., Shimizu, S., Ikenoue, T., Fukuma, S., Yamamoto, Y., Fukuhara, S.: Majority of systematic reviews published in high-impact journals neglected to register the protocols: a meta-epidemiological study. Journal of Clinical Epidemiology **84**, 54–60 (2017)

[119] Xu, C., Cheng, L.-L., Liu, Y., Jia, P.-L., Gao, M.-Y., Zhang, C.: Protocol registration or development may benefit the design, conduct and reporting of dose-response meta-analysis: empirical evidence from a literature survey. BMC medical research methodology **19**, 1–10 (2019)

[120] Polanin, J.R., Hennessy, E.A., Tsuji, S.: Transparency and reproducibility of meta-analyses in psychology: A meta-review. Perspectives on Psychological Science **15**(4), 1026–1041 (2020)

[121] So, T.: Classroom experiments as a replication device. Journal of Behavioral and Experimental Economics **86**, 101525 (2020)

[122] Ebersole, C.R., Atherton, O.E., Belanger, A.L., Skulborstad, H.M., Allen, J.M., Banks, J.B., Baranski, E., Bernstein, M.J., Bonfiglio, D.B., Boucher, L., *et al.*: Many labs 3: Evaluating participant pool quality across the academic semester via replication. Journal of Experimental Social Psychology **67**, 68–82 (2016)

[123] Klein, R., Ratliff, K., Vianello, M., Adams, R., Bahník, Š., Bernstein, M., Bocian, K., Brandt, M., Brooks, B., Brumbaugh, C., et al.: Investigating variation in replicability: A "many labs" replication project (2022)

[124] Glöckner, A., Jekel, M., Allstadt Torras, R., Dorrough, A.R., Anderl, C., Franke, N., Mischkowski, D., Fiedler, S., Miketta, S., Goltermann, J.: Hagen Cumulative Science Project. Project overview at osf.io/d7za8 (2015)

[125] Moshontz, H., Campbell, L., Ebersole, C.R., IJzerman, H., Urry, H.L., Forscher, P.S., Grahe, J.E., McCarthy, R.J., Musser, E.D., Antfolk, J., *et al.*: The psychological science accelerator: Advancing psychology through a distributed collaborative network. Advances in Methods and Practices in Psychological Science **1**(4), 501–515 (2018)

[126] Forscher, P.S., Wagenmakers, E.-J., Coles, N.A., Silan, M.A., Dutra, N., Basnight-Brown, D., IJzerman, H.: The benefits, barriers, and risks of big-team science. Perspectives on Psychological Science, 17456916221082970 (2020)

[127] Lieck, D.S.N., Lakens, D.: An Overview of Team Science Projects in the Social Behavioral Sciences. Project at http://doi.org/10.17605/OSF.IO/WX4ZD (2022)

[128] Jarke, H., Anand-Vembar, S., Alzahawi, S., Andersen, T.L., Bojanić, L., Carstensen, A., Feldman, G., Garcia-Garzon, E., Kapoor, H., Lewis, S., *et al.*: A roadmap to large-scale multi-country replications in psychology. Collabra: Psychology **8**(1), 57538 (2022)

[129] Pennington, C.R., Jones, A.J., Tzavella, L., Chambers, C.D., Button, K.S.: Beyond online participant crowdsourcing: The benefits and opportunities of big team addiction science. Experimental and Clinical Psychopharmacology (2022)

[130] Disis, M.L., Slattery, J.T.: The road we must take: multidisciplinary team science. Science translational medicine **2**(22), 22–9229 (2010)

[131] Ledgerwood, A., Hudson, S.-k.T.J., Lewis Jr, N.A., Maddox, K.B., Pickett, C.L., Remedios, J.D., Cheryan, S., Diekman, A.B., Dutra, N.B., Goh, J.X., *et al.*: The pandemic as a portal: Reimagining psychological science as truly open and inclusive. Perspectives on Psychological Science **17**(4), 937–959 (2022)

[132] Coles, N.A., March, D.S., Marmolejo-Ramos, F., Larsen, J.T., Arinze, N.C., Ndukaihe, I.L., Willis, M.L., Foroni, F., Reggev, N., Mokady, A., et al.: A multi-lab test of the facial feedback hypothesis by the many smiles collaboration. Nature human behaviour, 1–12 (2022)

[133] Coles, N.A., Gaertner, L., Frohlich, B., Larsen, J.T., Basnight-Brown, D.M.: Fact or artifact? demand characteristics and participants' beliefs can moderate, but do not fully account for, the effects of facial feedback on emotional experience. Journal of Personality and Social Psychology **124**(2), 287 (2023)

[134] Cowan, N., Belletier, C., Doherty, J.M., Jaroslawska, A.J., Rhodes, S., Forsberg, A., Naveh-Benjamin, M., Barrouillet, P., Camos, V., Logie, R.H.: How do scientific views change? notes from an extended adversarial collaboration. Perspectives on Psychological Science **15**(4), 1011–1025 (2020)

[135] Forscher, P.S., Taylor, V.J., Cavagnaro, D., Lewis Jr, N., Buchanan, E.M., Moshontz, H., Mark, A.Y., Appleby, S., Batres, C., Bennett-Day, B., et al.: Stereotype threat in black college students across many operationalizations (2019)

[136] Wang, K., Goldenberg, A., Dorison, C.A., Miller, J.K., Uusberg, A., Lerner, J.S., Gross, J.J., Agesin, B.B., Bernardo, M., Campos, O., *et al.*: A multi-country test of brief reappraisal interventions on emotions during the covid-19 pandemic. Nature Human Behaviour **5**(8), 1089–1110 (2021)

[137] Kahneman, D., Klein, G.: Conditions for intuitive expertise: a failure to disagree. American psychologist **64**(6), 515 (2009)

[138] Kekecs, Z., Palfi, B., Szaszi, B., Szecsi, P., Zrubka, M., Kovacs, M., Bakos, B.E., Cousineau, D., Tressoldi, P., Schmidt, K., *et al.*: Raising the value of research studies in psychological science by increasing the credibility of research reports: the transparent psi project. Royal Society Open Science **10**(2), 191375 (2023)

[139] Azevedo, F.: Not so simple: science is in the details. Psychological Inquiry **31**(1), 61–65 (2020)

[140] Henrich, J., Heine, S.J., Norenzayan, A.: The weirdest people in the world? Behavioral and brain sciences **33**(2-3), 61–83 (2010)

[141] Yarkoni, T.: The generalizability crisis. Behavioral and Brain Sciences **45**, 1 (2022)

[142] Ghai, S.: It's time to reimagine sample diversity and retire the weird dichotomy. Nature Human Behaviour **5**(8), 971–972 (2021)

[143] Nielsen, M.W., Andersen, J.P.: Global citation inequality is on the rise. Proceedings of the National Academy of Sciences **118**(7), 2012208118 (2021)

[144] Oberauer, K., Lewandowsky, S.: Addressing the theory crisis in psychology. Psychonomic bulletin & review **26**, 1596–1618 (2019)

[145] Scheel, A.M., Tiokhin, L., Isager, P.M., Lakens, D.: Why hypothesis testers should spend less time testing hypotheses. Perspectives on Psychological Science **16**(4), 744–755 (2021)

[146] Chauvette, A., Schick-Makaroff, K., Molzahn, A.E.: Open data in qualitative research. International Journal of Qualitative Methods **18**, 1609406918823863 (2019)

[147] Branney, P., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C.V., Treharne, G.J., Whitaker, C.: Three steps to open science for qualitative research in psychology (2022)

[148] Çelik, H., Baykal, N.B., Memur, H.N.K.: Qualitative data analysis and fundamental principles. Journal of Qualitative Research in Education **8**(1) (2020)

[149] Class, B., Bruyne, M., Wuillemin, C., Donzé, D., Claivaz, J.-B.: Towards open science for the qualitative researcher: From a positivist to an open interpretation. International Journal of Qualitative Methods **20**, 16094069211034641 (2021)

[150] Humphreys, L., Lewis Jr, N.A., Sender, K., Won, A.S.: Integrating qualitative methods and open science: Five principles for more trustworthy research. Journal of Communication **71**(5), 855–874 (2021)

[151] Steinhardt, I., Bauer, M., Wünsche, H., Schimmler, S.: The connection of open science practices and the methodological approach of researchers. Quality & Quantity, 1–16 (2022)

[152] Haven, T.L., Van Grootel, L.: Preregistering qualitative research. Accountability in research **26**(3), 229–244 (2019)

[153] Bergmann, C.: How to integrate open science into language acquisition research? Student workshop held at the 43rd Annual Boston University Conference on Language Development (BUCLD 43) (2018)

[154] Petersen, O.H.: Inequality of research funding between different countries and regions is a serious problem for global science. Function **2**(6), 060 (2021)

[155] Puthillam, A., Doble, L.J.M., Santos, J.J.D., Elsherif, M., Steltenpohl, C.N., Moreau, D., Pownall, M., Kapoor, H.: Guidelines to improve internationalization in psychological science (2022)

[156] Taffe, M., Gilpin, N.: Equity, diversity and inclusion: Racial inequity in grant funding from the US National Institutes of Health. eLife, 10, Article e65697 (2021)

[157] Burns, K.E., Straus, S.E., Liu, K., Rizvi, L., Guyatt, G.: Gender differences in grant and personnel award funding rates at the canadian institutes of health research based on research content area: A retrospective analysis. PLoS medicine **16**(10), 1002935 (2019)

[158] Sato, S., Gygax, P.M., Randall, J., Schmid Mast, M.: The leaky pipeline in research grant peer review and funding decisions: challenges and future directions. Higher Education **82**(1), 145–162 (2021)

[159] Guttinger, S.: The limits of replicability. European Journal for Philosophy of Science **10**(2), 10 (2020)

[160] Evans, T.: Developments in open data norms. Journal of Open Psychology Data **10**(1) (2022)

[161] Holcombe, A.O., Kovacs, M., Aust, F., Aczel, B.: Documenting contributions to scholarly articles using credit and tenzing. PLoS One **15**(12), 0244611 (2020)

[162] John, L.K., Loewenstein, G., Prelec, D.: Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological science **23**(5), 524–532 (2012)

[163] Simmons, J.P., Nelson, L.D., Simonsohn, U.: False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. (2016)

[164] Wicherts, J.M., Veldkamp, C.L., Augusteijn, H.E., Bakker, M., Van Aert, R., Van Assen, M.A.: Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. Frontiers in psychology, 1832 (2016)

[165] Agnoli, F., Wicherts, J.M., Veldkamp, C.L., Albiero, P., Cubelli, R.: Questionable research practices among italian research psychologists. PloS one **12**(3), 0172792 (2017)

[166] Fiedler, K., Schwarz, N.: Questionable research practices revisited. Social Psychological and Personality Science **7**(1), 45–52 (2016)

[167] Kerr, N.L.: Harking: Hypothesizing after the results are known. Personality and social psychology review **2**(3), 196–217 (1998)

[168] Flake, J.K., Pek, J., Hehman, E.: Construct validation in social and personality research: Current practice and recommendations. Social Psychological and Personality Science **8**(4), 370–378 (2017)

[169] Flake, J.K., Fried, E.I.: Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in

Psychological Science **3**(4), 456–465 (2020)

[170] Molléri, J.S.: Research incentives in academia leading to unethical behavior. In: Research Challenges in Information Science: 16th International Conference, RCIS 2022, Barcelona, Spain, May 17–20, 2022, Proceedings, pp. 744–751 (2022). Springer

[171] Gerrits, R.G., Jansen, T., Mulyanto, J., Berg, M.J., Klazinga, N.S., Kringos, D.S.: Occurrence and nature of questionable research practices in the reporting of messages and conclusions in international scientific health services research publications: a structured assessment of publications authored by researchers in the netherlands. BMJ open **9**(5), 027903 (2019)

[172] Checchi, D., De Fraja, G., Verzillo, S.: Incentives and careers in academia: theory and empirical analysis. The Review of Economics and Statistics **103**(4), 786–802 (2021)

[173] Grove, L.: The effects of funding policies on academic research. PhD thesis, University College London (2017)

[174] Frias-Navarro, D., Pascual-Soler, M., Perezgonzalez, J., Monterde-i-Bort, H., Pascual-Llobell, J.: Spanish scientists' opinion about science and researcher behavior. The Spanish Journal of Psychology **24**, 7 (2021)

[175] Bornmann, L., Daniel, H.-D.: The state of h index research: Is the h index the ideal way to measure research performance? EMBO reports **10**(1), 2–6 (2009)

[176] Primbs, M.A., Pennington, C.R., Lakens, D., Silan, M.A.A., Lieck, D.S., Forscher, P.S., Buchanan, E.M., Westwood, S.J.: Are small effects the indispensable foundation for a cumulative psychological science? a reply to götz et al.(2022). Perspectives on Psychological Science, 17456916221100420 (2022)

[177] Munafò, M., Noble, S., Browne, W.J., Brunner, D., Button, K., Ferreira, J., Holmans, P., Langbehn, D., Lewis, G., Lindquist, M., *et al.*: Scientific rigor and the art of motorcycle maintenance. Nature biotechnology **32**(9), 871–873 (2014)

[178] Martin, G., Clarke, R.M.: Are psychology journals anti-replication? a snapshot of editorial practices. Frontiers in Psychology **8**, 523 (2017)

[179] Cohen, B.A.: How should novelty be valued in science? Elife **6**, 28699 (2017)

[180] Tijdink, J.K., Vergouwen, A.C., Smulders, Y.M.: Publication pressure and burn out among dutch medical professors: A nationwide survey. PloS one **8**(9), 73381 (2013)

[181] Tijdink, J.K., Verbeke, R., Smulders, Y.M.: Publication pressure and scientific misconduct in medical scientists. Journal of Empirical Research on Human

Research Ethics **9**(5), 64–71 (2014)

[182] Laitin, D.D., Miguel, E., Alrababa'h, A., Bogdanoski, A., Grant, S., Hoeberling, K., Hyunjung Mo, C., Moore, D.A., Vazire, S., Weinstein, J., *et al.*: Reporting all results efficiently: A rare proposal to open up the file drawer. Proceedings of the National Academy of Sciences **118**(52), 2106178118 (2021)

[183] Scheel, A.M., Schijen, M.R., Lakens, D.: An excess of positive results: Comparing the standard psychology literature with registered reports. Advances in Methods and Practices in Psychological Science **4**(2), 25152459211007467 (2021)

[184] Franco, A., Malhotra, N., Simonovits, G.: Publication bias in the social sciences: Unlocking the file drawer. Science **345**(6203), 1502–1505 (2014)

[185] Matarese, V.: Kinds of replicability: Different terms and different functions. Axiomathes, 1–24 (2022)

[186] Maxwell, S.E., Lau, M.Y., Howard, G.S.: Is psychology suffering from a replication crisis? what does "failure to replicate" really mean? American Psychologist **70**(6), 487 (2015)

[187] Ulrich, R., Miller, J.: Questionable research practices may have little effect on replicability. Elife **9**, 58237 (2020)

[188] Stroebe, W., Strack, F.: The alleged crisis and the illusion of exact replication. Perspectives on Psychological Science **9**(1), 59–71 (2014)

[189] Feest, U.: Why replication is overrated. Philosophy of Science **86**(5), 895–905 (2019)

[190] Devezer, B., Navarro, D.J., Vandekerckhove, J., Ozge Buzbas, E.: The case for formal methodology in scientific reform. Royal Society open science **8**(3), 200805 (2021)

[191] Eronen, M.I., Bringmann, L.F.: The theory crisis in psychology: How to move forward. Perspectives on Psychological Science **16**(4), 779–788 (2021)

Table 1 Why research fails to replicate. This box outlines some explanations for the low replicability of research which we exemplify at the individual and structural level. A more exhaustive overview can be found in Pennington [15].

Term	***Definition***
Credibility Revolution	The problems and the solutions resulting from a growing distrust in scientific findings, following concerns about the credibility of scientific claims (e.g., low replicability). The term has been proposed as a more positive alternative to the term replicability crisis, and includes the many solutions to improve the credibility of research, such as pre-registration, transparency, and replication.
Open Science	An umbrella term reflecting the idea that scientific knowledge of all kinds, where appropriate, should be openly accessible, transparent, rigorous, reproducible, replicable, accumulative, and inclusive, all which are considered fundamental features of the scientific endeavour. Open science consists of principles and behaviours that promote transparent, credible, reproducible, and accessible science. Open science has six major aspects: open data, open methodology, open source, open access, open peer review, and open educational resources.
Open Scholarship	'Open scholarship' is often used synonymously with 'open science', but extends to all disciplines, drawing in those which might not traditionally identify as science-based. It reflects the idea that knowledge of all kinds should be openly shared, transparent, rigorous, reproducible, replicable, accumulative, and inclusive (allowing for all knowledge systems). Open scholarship includes all scholarly activities that are not solely limited to research such as teaching and pedagogy.
Questionable Research Practices	A range of activities that intentionally or unintentionally distort data in favour of a researcher's own hypotheses - or omissions in reporting such practices - including; selective inclusion of data, hypothesising after the results are known (HARKing), and p-hacking. Popularised by John et al. [162].
Replicability	An umbrella term, used differently across fields, covering concepts of: direct and conceptual replication, computational reproducibility/replicability, generalizability analysis and robustness analyses. Some of the definitions used previously include: a different team arriving at the same results using the original author's artefacts; a study arriving at the same conclusion after collecting new data; as well as studies for which any outcome would be considered diagnostic evidence about a claim from prior research.
Replication Crisis	The finding, and related shift in academic culture and thinking, that a large proportion of scientific studies published across disciplines do not replicate (e.g., [6]). This is considered to be due to a lack of quality and integrity of research and publication practices, such as publication bias, questionable research practices and a lack of transparency, leading to an inflated rate of false positive results. Others have described this process as a 'Credibility Revolution' towards improving these practices.
Reproducibility	A minimum standard on a spectrum of activities ("reproducibility spectrum") for assessing the value or accuracy of scientific claims based on the original methods, data, and code. For instance, where the original researcher's data and computer codes are used to regenerate the results, often referred to as computational reproducibility. Reproducibility does not guarantee the quality, correctness, or validity of the published results. In some fields, this meaning is, instead, associated with the term "replicability" or 'repeatability'.
Transparency	Transparency refers to a combination of availability and accountability, or practically, having one's actions open and accessible for external evaluation. Transparency pertains to researchers being honest about theoretical, methodological, and analytical decisions made throughout the research cycle. Transparency can be usefully differentiated into "scientifically relevant transparency" and "socially relevant transparency". While the former has been the focus of early Open Science discourses, the latter is needed to provide scientific information in ways that are relevant to decision makers and members of the public.

Box 1. Why research fails to replicate. This box outlines some explanations for the low replicability of research which we exemplify at the individual and structural level. A more exhaustive overview can be found in [15].

Individual level. Most studies seeking to uncover causes of the replicability crisis have, to date, heavily focused on questionable research practices (QRPs) at the level of individual researchers. For instance, research has highlighted how researchers have many 'degrees of freedom' when processing and analysing data, such as excluding outliers or choosing to run multiple tests on subsets of the data, and the impact this has on false positive rates [163, 164]. Such flexibility fails to prevent researchers from employing a variety of QRPs, which have been repeatedly cited as substantial contributors to low replicability [162, 164–167]. QRPs can also extend to distorting measurement (e.g., omitting to report the reliability or validity of an instrument when found to be unsatisfactory; [168, 169]), likely with similar effects for replicability. Finally, a lack of transparency in the reporting can mask and/or exacerbate these issues, contributing to poor replicability and reproducibility of research [15].

Structural level. Characteristics inherent in the academic system also contribute widely to low replicability by setting challenging and in itself questionable research goals. This focus can have downstream effects for researchers trying to achieve these goals to obtain career stability or advancement, and encourage the usage of unethical research behaviours [170] such as QRPs [84, 171]. More generally, the incentivisation of research is conveyed through common academic aphorisms such as 'the winner takes it all' or 'publish or perish' [172]. Many of these incentives are not only pushed by research institutions and their governing agencies but are also influenced by publishers and funding agencies [173]. For example, the emphasis on arbitrary publication metrics, such as the impact factor [174] or the h-index contribute to the replication crisis by creating perverse incentives [175]. Such emphasis is particularly problematic with regards to publication practices, where novel and hypothesis-supporting research is viewed more positively by editors and reviewers, and thus published more frequently [176] and when research quantity is prioritised over quality [36, 74, 177], while there were historically little-to-no requirements for transparency in the publishing process [160]. Consequently, most journals have prevented or resisted publication of null-findings and replications [178] often due to perceived 'lack of contribution', or selecting novelty over previously researched [179–181]. This selective publishing practice creates 'publication bias', the distortion of the literature to over-represent positive findings and under-represent negative ones, giving distorted or misleading representations of existing effects [46, 182, 183]. Negative findings instead land in the metaphorical 'file-drawer' and are often never published at all [184].

Challenges with replications. Replicability is a highly heterogeneous concept [185], not applying equally to all research fields, as for example the standardisation and control for covariates vary for different objects of research [159]. Hence, replications are not the silver bullet to ensure scientific quality. Differences in the replicability of research are often attributed to features of both original studies and their replications, such as procedures, measurement characteristics or evidence strength [8, 186, 187]. Other relevant features are the direct characteristics of the effects themselves, which may vary in size depending on time or context [188]. However, low reliability and systematic error in original studies [189] or missing formalisations of verbalisations, for example of concepts, definitions, and results [190], can lead to mismatches between used statistical tests and their interpretation [141]. Furthermore, inappropriate use of statistical tests [83, 84] and wrong interpretation of such [141], as well as weak theoretical development underpinning hypothesised effects [191] or a skewed baseline set of studies due to publication bias [46, 145, 176] likely influence replication rates.