

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

Max Korbmacher<sup>1,2,3</sup>, Flavio Azevedo<sup>4,35</sup>✉, Charlotte R. Pennington<sup>5</sup>,  
Helena Hartmann<sup>6</sup>, Madeleine Pownall<sup>7</sup>, Kathleen Schmidt<sup>8</sup>,  
Mahmoud Elsherif<sup>9</sup>, Nate Breznau<sup>10</sup>, Olly Robertson<sup>11</sup>, Tamara Kalandadze<sup>12</sup>,  
Shijun Yu<sup>9</sup>, Bradley J. Baker<sup>13</sup>, Aoife O'Mahony<sup>14</sup>, Jørgen Ø. -S. Olsnes<sup>15</sup>,  
John J. Shaw<sup>16</sup>, Biljana Gjonjeska<sup>17</sup>, Yuki Yamada<sup>18</sup>, Jan P. Röer<sup>19</sup>,  
Jennifer Murphy<sup>20</sup>, Shilaan Alzahawi<sup>21</sup>, Sandra Grinschgl<sup>22</sup>,  
Catia M. Oliveira<sup>23</sup>, Tobias Wingen<sup>24</sup>, Siu Kit Yeung<sup>25</sup>, Meng Liu<sup>26</sup>,  
Laura M. König<sup>27</sup>, Nihan Albayrak-Aydemir<sup>28,29</sup>, Oscar Lecuona<sup>30,31</sup>,  
Leticia Micheli<sup>32</sup> & Thomas Evans<sup>33,34</sup>

The emergence of large-scale replication projects yielding successful rates substantially lower than expected caused the behavioural, cognitive, and social sciences to experience a so-called 'replication crisis'. In this Perspective, 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 on our research environment.

**A**fter several notable controversies in 2011<sup>1–3</sup>, skepticism regarding claims in psychological science increased and inspired the development of projects examining the replicability and reproducibility of past findings<sup>4</sup>. Replication refers to the process of repeating a study or experiment with the goal of verifying effects or generalising findings across new models

<sup>1</sup> Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway. <sup>2</sup> NORMENT Centre for Psychosis Research, University of Oslo and Oslo University Hospital, Oslo, Norway. <sup>3</sup> Mohn Medical Imaging and Visualisation Center, Bergen, Norway. <sup>4</sup> Department of Psychology, University of Cambridge, Cambridge, UK. <sup>5</sup> School of Psychology, Aston University, Birmingham, UK. <sup>6</sup> Department of Neurology, University of Essen, Essen, Germany. <sup>7</sup> School of Psychology, University of Leeds, Leeds, UK. <sup>8</sup> Department of Psychology, Ashland University, Ashland, USA. <sup>9</sup> School of Psychology, University of Birmingham, Birmingham, UK. <sup>10</sup> SOCIUM Research Center on Inequality and Social Policy, University of Bremen, Bremen, Germany. <sup>11</sup> Department of Psychiatry, University of Oxford, Oxford, UK. <sup>12</sup> Department of Education, ICT and Learning, Ostfold University College, Halden, Norway. <sup>13</sup> Department of Sport and Recreation Management, Temple University, Philadelphia, USA. <sup>14</sup> School of Psychology, Cardiff University, Cardiff, UK. <sup>15</sup> Kavli Institute for Systems Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway. <sup>16</sup> Division of Psychology, De Montfort University, Leicester, UK. <sup>17</sup> Macedonian Academy of Sciences and Arts, Skopje, North Macedonia. <sup>18</sup> Faculty of Arts and Science, Kyushu University, Fukuoka, Japan. <sup>19</sup> Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany. <sup>20</sup> Department of Applied Science, Technological University Dublin, Dublin, Ireland. <sup>21</sup> Graduate School of Business, Stanford University, Stanford, USA. <sup>22</sup> Institute of Psychology, University of Graz, Graz, Austria. <sup>23</sup> Department of Psychology, University of York, York, UK. <sup>24</sup> Institute of General Practice and Family Medicine, University of Bonn, Bonn, Germany. <sup>25</sup> Department of Psychology, Chinese University of Hong Kong, Hong Kong, China. <sup>26</sup> Faculty of Education, University of Cambridge, Cambridge, UK. <sup>27</sup> Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Bayreuth, Germany. <sup>28</sup> Open Psychology Research Centre, Open University, Milton Keynes, UK. <sup>29</sup> Department of Psychological and Behavioural Science, London School of Economics and Political Science, London, UK. <sup>30</sup> Department of Psychology, Universidad Rey Juan Carlos, Madrid, Spain. <sup>31</sup> Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain. <sup>32</sup> Institute of Psychology, Leiden University, Leiden, The Netherlands. <sup>33</sup> School of Human Sciences, University of Greenwich, Greenwich, UK. <sup>34</sup> Institute for Lifecourse Development, University of Greenwich, Greenwich, UK. <sup>35</sup> Present address: Department of Social Psychology, University of Groningen, Groningen, The Netherlands. ✉email: [flavio.azevedo@rug.nl](mailto:flavio.azevedo@rug.nl)

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).

In one of the most impactful replication initiative of the last decade, the Open Science Collaboration<sup>5</sup> 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 a wider context, a minimum replicability rate of 89% should have been expected if all of the original effects were true (and not false positives; ref. 6). Pooling the Open Science Collaboration<sup>5</sup> replications with 207 other replications from recent years resulted in a higher estimate; 64% of effects successfully replicated with effect sizes being 32% smaller than the original effects<sup>7</sup>. While estimations of replicability may vary, they nevertheless appear to be sub-optimal—an issue that is not exclusive to psychology and found across many other disciplines (e.g., animal behaviour<sup>8–10</sup>; cancer biology<sup>11</sup>; economics<sup>12</sup>), and symptomatic of persistent issues within the research environment<sup>13,14</sup>. The ‘replication crisis’ has introduced a number of considerable challenges, including compromising the public’s trust in science<sup>15</sup> and undermining the role of science and scientists as reliable sources to inform evidence-based policy and practice<sup>16</sup>. At the same time, the crisis has

provided a unique opportunity for scientific development and reform. In this narrative review, we focus on the latter, exploring the replication crisis through the lens of a credibility revolution<sup>17</sup> to provide an overview of recent developments that have led to positive changes in the research landscape (see Fig. 1).

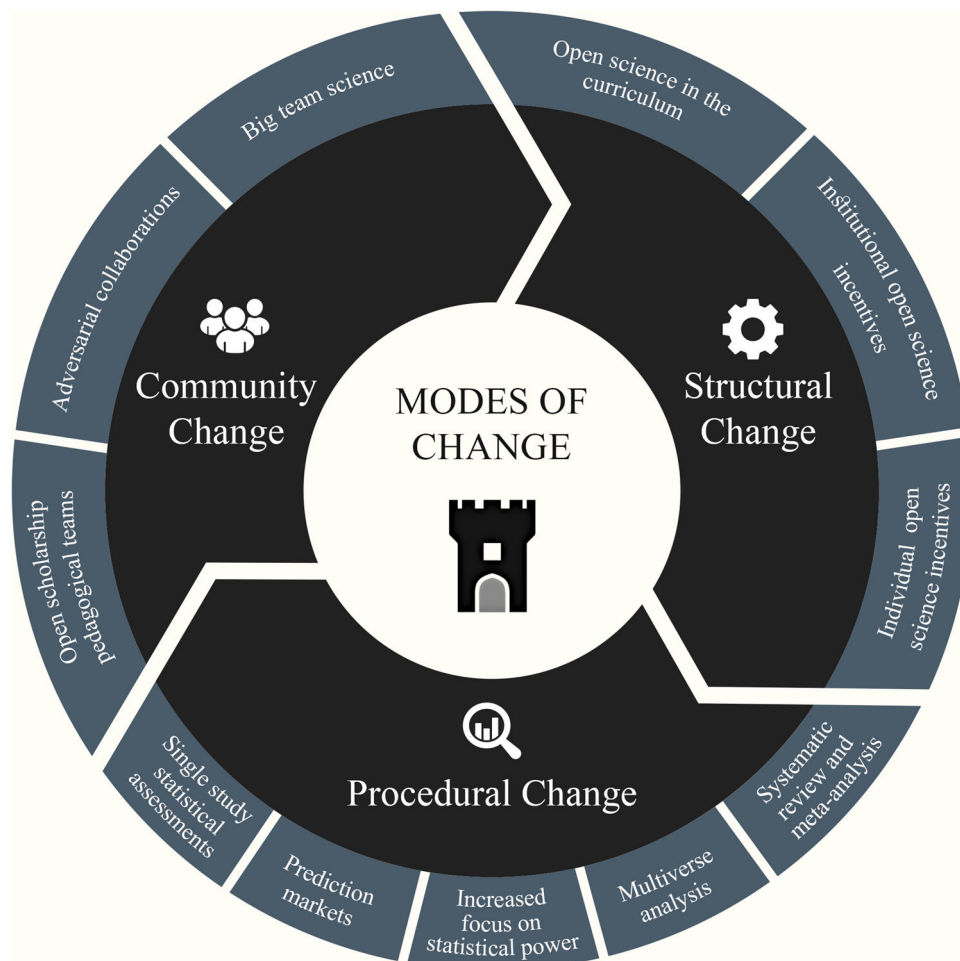
Recent discussions have outlined various reasons why replications fail (see Box 1). To address these replicability concerns, different perspectives have been offered on how to reform and promote improvements to existing research norms in psychological science<sup>18–20</sup>. An academic movement collectively known as open scholarship (incorporating Open Science and Open Research) has driven constructive change by accelerating the uptake of robust research practices while concomitantly championing a more diverse, equitable, inclusive, and accessible psychological science<sup>21,22</sup>.

These reforms have been driven by a diverse range of institutional initiatives, grass-roots, bottom-up initiatives, and individuals. The extent of a such impact led Vazire<sup>17</sup> to reframe the replicability crisis as a credibility revolution, acknowledging that 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.

Scientific practices are behaviours<sup>23</sup> and can be changed, especially when structures (e.g., funding agencies), environments (e.g., research groups), and peers (e.g., individual researchers)

**Table 1 Terminology adopted from the glossary of Open Science terminology<sup>50</sup>.**

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 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. <sup>202</sup> .
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., ref. 5). 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, starting with Vazire <sup>17</sup> , 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.



**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 and expressed by new norms and rules; procedural change refers to behaviours and sets of commonly used practices in the research process; community change encompasses how work and collaboration within the scientific community evolves.

facilitate and support them. Most attempts to change the behaviours of individual researchers have concentrated on identifying and eliminating problematic practices and improving training in open scholarship<sup>23</sup>. Efforts to change individuals' behaviours have ranged from the creation of grass-roots communities to support individuals to incorporate open scholarship practices into their research and teaching (e.g., ref. <sup>24</sup>) to infrastructural change (e.g., creation of open tools fostering the uptake of improved norms such as the software StatCheck<sup>25</sup> to identify statistical inconsistencies en-masse, providing high-quality and modularized training on the underlying skills needed for transparent and reproducible data preparation and analysis<sup>26</sup> or documenting contributions and author roles transparently<sup>27</sup>).

The replication crisis has highlighted the need for a deeper understanding of the research landscape and culture, and a concerted effort from institutions, funders, and publishers to address the substantive issues. Despite the creation of new open access journals, they still face challenges in gaining acceptance due to the prevailing reputation and prestige-based market. These stakeholders have made significant efforts, but their impact remains isolated and infrequently reckoned. As a result, although there have been positive developments, progress toward a systemic transformation in how science is considered, actioned, and structured is still in its infancy.

In this article, we take the opportunity to reflect upon the scope and extent of positive changes resulting from 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 optimistic changes allowing us to empower and retain change-makers towards further scientific reform.

### Structural change

In the wake of the credibility revolution, structural change is seen as crucial to achieving the goals of open scholarship, with new norms and rules often being developed at the institutional level. In this context, there has been increasing interest in embedding open scholarship practices into the curriculum and incentivizing researchers to adopt improved practices. In the following, we describe and discuss examples of structural change and its impact.

**Embedding replications into the curriculum.** Higher Education instructors and programmes have begun integrating 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<sup>28</sup>, and there are freely

**Box 1 | Why research fails to replicate. This box outlines some explanations for the low replicability of research at the individual and structural levels. A more exhaustive overview can be found in ref. 14**

*Individual level.* Questionable research practices (QRPs) at the individual researcher level may help understand the replicability crisis. Researchers have significant flexibility in processing and analysing data, including the exclusion of outliers and running multiple tests on subsets of the data, leading to false positive results<sup>203,204</sup>. QRPs also involve measurement, such as omitting to report psychometric properties when found to be unsatisfactory, potentially compromising replicability<sup>205,206</sup>, likely with similar negative effects for replicability. Researchers continue to employ a variety of QRPs, which are significant contributors to low replicability<sup>202,204,207–209</sup>, and a lack of transparency in reporting can mask and exacerbate these issues<sup>14</sup>. *Structural level.* Characteristics of the academic system also contribute widely to low replicability. For example, misaligned incentives can influence researchers trying to obtain career stability or advancement and encourage the usage of unethical research behaviours<sup>210</sup> such as QRPs<sup>103,211</sup>. More generally, the incentivisation of research is conveyed through common academic aphorisms such as ‘publish or perish’<sup>212</sup>. Many of these incentives are driven by research institutions and their governing agencies but also by publishers and funding agencies<sup>213</sup>. For example, the emphasis on arbitrary publication metrics, such as the impact factor<sup>214</sup> or the h-index, create perverse incentives<sup>215</sup> that ultimately reinforce the prioritization of research quantity over quality<sup>85,92,216</sup>, with little-to-no requirements for transparency in the publishing process<sup>201</sup>. Such emphasis is particularly evident in publication practices, where novel and hypothesis-supporting research has historically been viewed more positively by editors and reviewers, and thus published more frequently<sup>217</sup>. Many journals prevent or resist publication of null-findings and replications<sup>218</sup> often due to perceived ‘lack of contribution’, or prioritising novelty over incremental developments<sup>219–221</sup>. This selective publishing creates ‘publication bias’, a distortion of the literature to over-represent positive findings and under-represent negative ones, giving misleading representations of existing effects<sup>50,78,222</sup>. Negative findings can often end in the ‘file-drawer’ and are never published<sup>223</sup>.

*Challenges with replications.* While replicability is an essential feature of scientific research, its application can vary depending on the research field, with standardization and control of covariates varying for different objects of study<sup>200</sup>. Moreover, replication is not conclusive in ensuring scientific quality due to the heterogeneity of the concept<sup>224</sup>. Differences in replicability can be attributed to several factors such as variations in procedures, measurement characteristics, or evidence strength between original studies and their replications<sup>7,225–227</sup>. Additionally, the characteristics of the effects themselves may vary in size depending on time or context<sup>228</sup>. However, low reliability and systematic errors in original studies<sup>229</sup>, missing formalizations of verbalizations, for example of concepts, definitions, and results<sup>174</sup>, or the inappropriate use and interpretation of statistical tests<sup>102,103,170</sup>, and weak theoretical development underpinning hypothesized effects<sup>230</sup> can lead to mismatches between statistical tests and their interpretations, ultimately influencing replication rates.

available, curated materials covering the entire process of executing replications with students<sup>29</sup> (see also [forrt.org/reversals](http://forrt.org/reversals)). In one prominent approach, the Collaborative Replications and Education Project<sup>30,31</sup> 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<sup>32–34</sup>, graduate students complete replication projects as part of their dissertations. Early career researchers (ECRs) are invited to prepare the manuscripts for publication<sup>35,36</sup> and, in this way, students’ research efforts for their dissertation are utilised to contribute to a more robust body of literature, while being formally acknowledged. An additional benefit is the opportunity for ECRs to further their career by publishing available data. Institutions and departments can also profit from embedding these projects as these not only increase the quality of education on research practices and transferable skills but also boost research outputs<sup>21,37</sup>.

If these models are to become commonplace, developing a set of standards regarding authorship is 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; ref. 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<sup>27,39,40</sup>.

**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 toward becoming more effective and responsible researchers and consumers of science<sup>41</sup>. It is therefore essential to adequately address open scholarship in the classroom, and promote the creation and maintenance of open educational resources supporting teaching staff<sup>21,29,37</sup>. Gaining an increased scientific literacy early on may have significant long-term benefits for students, including the opportunity to make a rigorous scientific contribution, acquire a critical understanding of the scientific process and the value of replication, and a commitment to research integrity values such as openness and transparency<sup>24,41–43</sup>. 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<sup>44</sup>. This creates a path towards open scholarship values and practices becoming the norm rather than the exception. It also links directly to existing social movements often embraced among university students to foster greater equity and justice and helps break down status hierarchies and power structures (e.g., decolonisation, diversity, equity, inclusion and accessibility efforts)<sup>45–49</sup>.

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 rigour and transparency over time. While these changes are structural, they are often driven by single or small groups of individuals, who are usually in the early stages of their careers and receive little recognition for their contributions<sup>21</sup>. An increasing number of grassroots open science organisations contribute to different educational roles and provide resources, guidelines, and community. 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<sup>24</sup>. The FORRT community is currently running more than 15 initiatives which include summaries of open scholarship literature, a crowdsourced glossary of open scholarship terms<sup>50</sup>, a

literature review of the impact on students of integrating open scholarship into teaching<sup>41</sup>, out-of-the-box lesson plans<sup>51</sup>, a team working on bridging neurodiversity and open scholarship<sup>46,52</sup>, and a living database of replications and reversals<sup>53</sup>. Other examples of organisations providing open scholarship materials are ReproducibiliTea<sup>54</sup>, the RIOT Science Club, the Turing Way<sup>55</sup>, Open Life Science<sup>56</sup>, OpenScience<sup>57</sup>, the Network of Open Science Initiatives<sup>58</sup>, Course Syllabi for Open and Reproducible Methods<sup>59</sup>, the Carpentries<sup>60</sup>, the Embassy of Good Science<sup>61</sup>, the Berkeley Initiative for Transparency in Social Sciences<sup>62</sup>, the Institute for Replication<sup>63</sup>, Reproducibility for Everyone<sup>64</sup>, the International Network of Open Science<sup>65</sup>, and Reproducibility Networks such as the UKRN<sup>66</sup>.

These and other collections of open-source teaching and learning materials (such as podcasts, how-to guides, courses, labs, networks, and databases) can facilitate the integration of open scholarship principles into education and practice. Such initiatives not only raise awareness for open scholarship but also level the playing field for researchers from countries or institutions with fewer resources, such as the Global South and low- and middle-income countries, referring to the regions outside of Western Europe and North America that are primarily politically and culturally marginalised such as regions in Asia, Latin America, and Africa<sup>51,67</sup>.

**Incentives.** Scientific practice has been characterized by problematic rewards, such as prioritizing research quantity over quality and emphasizing statistically significant results. To foster a sustained integration of open scholarship practices, it is essential to revise incentive structures. Current efforts have focused on developing incentives that target various actors, including students, academics, faculties, universities, funders, and journals<sup>68–70</sup>. However, as each of these actors has different—and sometimes competing—goals, their motivations to engage in open scholarship practices can vary. In the following, we discuss recently developed incentives that specifically target researchers, academic journals, and funders.

*Targeting researchers.* Traditional incentives for academics to advance in their career are publishing articles, winning grants, and signalling the quality of the published work (e.g., perceived journal prestige)<sup>45,71</sup>. In some journals, researchers are now given direct incentives for the preregistration of study plans and analyses before study execution, and for openly sharing data and materials in the form of (open science) badges, with the aim of signalling study quality<sup>23</sup>. However, the extent to which badges can be used to increase open scholarship behaviours remains unclear; while one study<sup>72</sup> reports increased data sharing rates among articles published in Psychological Science with badges, a recent randomized control trial shows no evidence for badges increasing data sharing<sup>73</sup>, suggesting that more effective incentives or complementary workflows are required to motivate researchers to engage in open research practices<sup>74</sup>.

Furthermore, there are incentives provided for different open scholarship practices, such as using the Registered Report publishing format<sup>75,76</sup>. Here, authors submit research protocols for peer-review before data collection or analyses (in the case of secondary data). Registered Reports meeting high scientific standards are given provisional acceptance (‘in-principle acceptance’) before the results are known. Such format shifts the focus from the research outcomes to methodological quality and realigns incentives by providing researchers with the certainty of publication when adhering to the preregistered protocol<sup>75,76</sup>. 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<sup>77</sup>, while also allowing to report more negative results<sup>78</sup>, which may provide further incentives for researchers to adopt this format.

*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 open science standards to remain current and competitive by significantly increasing open publishing options and formulating new guidelines reinforcing and enforcing these changes<sup>79</sup>. For example, the Center for Open Science introduced the Transparency and Openness Promotion (TOP) Guidelines<sup>80</sup> comprising 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. Based on TOP and other Open Science best-practices, there are also guides for editors available (e.g., ref. <sup>81</sup>). Similarly, organisations such as NASA<sup>82</sup>, UNESCO<sup>83</sup>, and the European Commission<sup>84</sup> all came to support open scholarship efforts publicly, and on an international level. There are moreover efforts to open up funding options through the Registered Reports funding schemes<sup>85</sup>. Here, funding allocation and publication review are being combined into a single process, reducing both 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<sup>86,87</sup>, mandating open publishing when the research is funded by public grants. The increase in open access options illustrates how journals are being effectively incentivized to expand their repertoire and normalize open access<sup>88</sup>. At the same time, article processing charges have increased, causing new forms of inequities between the haves and have-nots, and the exclusion of researchers from low-resource universities across the globe<sup>88,89</sup>. As Plan-S shapes the decision space of journals and researchers, it is an incentive with the promise of long-term change.

Several initiatives aim to re-design systems such as peer review and publishing. Community peer reviews (e.g., PeerCommunityIn<sup>90</sup>) 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 shown by the recent editorial plans in the journal e-Life to abolish accept/reject decisions during peer review<sup>91</sup>, and as reflected by a recommendation-based system of the community peer review system.

Evaluation of researchers in academic settings has historically been focused on the quantity of papers they publish in high-impact journals<sup>45</sup>, despite criticism of the impact factor metric<sup>71</sup>, and their ability to secure grants<sup>85,92</sup>. In response to this narrow evaluation, a growing number of research stakeholders, including universities, have signed declarations such as the San Francisco Declaration of Research Assessment (DORA) or the agreement by the Coalition for Advancing Research Assessment (CORA). This initiative aims to broaden the criteria used for hiring, promotion, and funding decisions by considering all research outputs, including software and data, and considering the qualitative impact of research, such as its policy or practice implications. To promote sustained change towards open scholarship practices, some institutions have also modified the requirements for hiring

committees to consider such practices<sup>69</sup> (see, for example, the Open Hiring Initiative<sup>93</sup>). Such initiatives incentivize researchers to adopt open scholarship principles for career advancement.

### Procedural change

Procedural change refers to 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.

**Prediction markets of research credibility.** In recent years, researchers have employed prediction markets to assess the credibility of research findings<sup>94–99</sup>. 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%<sup>96,97</sup>). 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<sup>100</sup>. Replication markets are considered both an alternative and complementary approach to replication since certain conditions may favour one approach over the other. For instance, replication markets may be advantageous in cases where data collection is resource-intensive but less so when study design is especially complex. Therefore, replication markets offer yet another tool for researchers to assess the credibility of existing and hypothetical works. In that sense, it is an ongoing discussion whether low credibility estimates from replication markets can be used to inform decisions on which articles to replicate<sup>101</sup>.

**Statistical assessment tools.** Failure to control error rates and design high-power studies can contribute to low replication rates<sup>102,103</sup>. In response, researchers have developed various quantitative methods to assess expected distributions of statistical estimates (i.e.,  $p$ -values), such as  $p$ -curving<sup>104</sup>,  $z$ -curving<sup>105</sup>, and others.  $P$ -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<sup>104</sup>. Like  $p$ -curve, the  $z$ -curve assesses the distribution of test statistics while considering the power of statistical tests and false discovery rate within a body of literature<sup>105</sup>. 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  $\alpha$ -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  $\alpha$ -levels for all tests (e.g., from 0.05 to 0.005<sup>106</sup>), rethinking null hypothesis statistical testing (NHST) and considering exploratory NHST<sup>107</sup>, or abandoning NHST altogether<sup>108</sup> (see for an example, ref. <sup>109</sup>). However, these approaches are not panaceas and are unlikely to address all the highlighted concerns<sup>19,110</sup>. 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<sup>110</sup>. In this context, equivalence testing<sup>111</sup> or Bayesian analyses<sup>112</sup> have been proposed as suitable approaches to directly assess evidence for the alternative hypothesis against

evidence for the null hypothesis<sup>113</sup>. Graphical user interface (GUI) based statistical software packages, like JASP<sup>114</sup> and Jamovi<sup>115</sup>, have played a significant role in making statistical methods such as equivalence tests and Bayesian statistics accessible to a broader audience. The promotion of these methods, including practical walkthroughs and interactive tools like Shiny apps<sup>111,112</sup>, has further contributed to their increased adoption.

**Single-study statistical assessments.** A range of useful tools has 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<sup>25</sup>. Validation studies<sup>25</sup> reported high sensitivity (larger than 83%), specificity (larger than 96%), and accuracy (larger than 92%) of this tool. Other innovations include the Granularity-Related Inconsistency of Means (GRIM) test<sup>116</sup> aiming to evaluate the consistency of mean values of integer data (e.g., from Likert-type scales), considering sample size and the number of items. Another is the Sample Parameter Reconstruction via Iterative TEchniques (SPRITE), which reconstructs samples and estimates of the item value distributions based on reported descriptive statistics<sup>117</sup>. Adopting these efforts can serve as an initial step in reviewing existing literature, to ensure that findings are not the result of statistical errors or potential falsification. Researchers themselves can implement these tools to check their work and identify potential errors. With greater awareness and use of such tools, we can increase accessibility and enhance our ability to identify unsubstantiated claims.

**Multiverse analysis.** The multitude of researcher degrees of freedom—i.e., decisions researchers can make when using data—have been shown to influence the outcomes of analyses performed on the same data<sup>118,119</sup>. 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: no two groups used the same pipeline to pre-process the imaging data, which ultimately influenced both results and inference drawn from the results<sup>118</sup>. Another systematic effort comprising of 161 researchers in 73 teams independently investigated and tested a hypothesis central to an extensive body of scholarship using identical cross-country survey data—that more immigration will reduce public support for government provision of social policies—revealing a hidden universe of uncertainty<sup>120</sup>. The study highlights that the scientific process involves numerous analytical decisions that are often taken for granted as nondeliberate actions following established procedures but whose cumulative effect is far from insignificant<sup>119</sup>. It also illustrates that, even in a research setting with high accuracy motivation and unbiased incentives, reliability between researchers may remain low, regardless of researchers' methodological expertise<sup>119</sup>. The upshot is that idiosyncratic uncertainty may be a fundamental aspect of the scientific process that is not easily attributable to specific researcher characteristics or analytical decisions. However, increasing transparency regarding researcher degrees of freedom is still crucial, and multiverse analyses provide a useful tool to achieve this. By considering a range of feasible and reasonable analyses, researchers can test the same hypothesis across various scenarios and determine the stability of certain effects as they navigate the often large 'garden of forking paths'<sup>121</sup>. Multiverse analyses (and their sibling, sensitivity analyses) can now be performed more often to provide authoritative evidence of an effect on substantive research questions<sup>122–127</sup>. Such an approach helps researchers to determine the robustness of a finding by pooling evidence from a range of appropriate analyses.

**Systematic review and meta-analysis.** Systematic reviews or meta-analyses are used to synthesise findings from several primary studies<sup>128,129</sup>, 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<sup>130</sup>. 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<sup>131,132</sup>. Thus, researchers have developed best practices for open, and reproducible systematic reviews and meta-analyses such as Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)<sup>133</sup>, Non-Interventional, Reproducible, and Open (NIRO)<sup>134</sup> systematic review guidelines, the Generalized Systematic Review Registration Form<sup>135</sup>, and PROSPERO, a register of systematic review protocols<sup>136</sup>. These guides and resources provide opportunities for more systematic accounts of research<sup>137</sup>. These guidelines often include a risk of bias assessment, where different biases are assessed and reported<sup>138</sup>. 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<sup>139</sup>. An evaluated trial of enhanced requirements by some journals as part of the submission process<sup>140</sup> did lead to a slowly increasing uptake in such practices<sup>141</sup>, with later findings indicating that protocol registrations increased the quality of associated meta-analyses<sup>142</sup>. Optimistically, continuous efforts to increase transparency appear to have already contributed to researchers more consistently reporting eligibility criteria, effect size information, and synthesis techniques<sup>143</sup>.

### Community change

Community change encompasses how work and collaboration within the scientific community evolves. We describe two of these recent developments: Big Team Science and adversarial collaborations.

**Big Team Science.** The credibility revolution has undoubtedly driven the formation and development of various large-scale, collaborative communities<sup>144</sup>. Community examples of such approaches include mass replications, which can be integrated into research training<sup>30,31,34,145</sup>, and projects conducted by large teams and organisations such as the Many Labs studies<sup>146,147</sup>, the Hagen Cumulative Science Project<sup>148</sup>, the Psychological Science Accelerator<sup>149</sup>, and the Framework for Open and Reproducible Research Training (FORRT)<sup>24</sup>.

A promising development to accelerate scientific progress is Big Team Science—i.e., large-scale collaborations of scientists working on a scholarly common goal and pooling resources across labs, institutions, disciplines, cultures, and countries<sup>14,150,151</sup>. Replication studies are often the focus of such collaborations, with many of them sharing their procedural knowledge and scientific insights<sup>152</sup>. This collaborative approach leverages the expertise of a consortium of researchers, increases research efficiency by pooling resources such as time and funding, and allows for richer cross-cultural samples to draw conclusions from<sup>150,153</sup>. Big Team Science emphasizes various practices to improve research quality, including interdisciplinary internal reviews, incorporating multiple perspectives, implementing uniform protocols across participating labs, and recruiting larger and more diverse samples<sup>41,50,149,150,152,154</sup>. 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<sup>155</sup>.

Big-Team Science represents a prime opportunity to advance open scholarship goals that have proven to be the most difficult to achieve, including diversity, equity, inclusion, accessibility, and social justice in research. Through a collaborative approach that prioritizes the inclusion of disenfranchised researchers, coalition building, and the redistribution of expertise, training, and resources from the most to the least affluent, Big Team Science can contribute to a more transparent and robust science that is also more inclusive, diverse, accessible, and equitable. This participatory approach creates a better and more just science for everyone<sup>52</sup>. However, it is important to critically examine some of the norms, practices, and culture associated with Big Team Science to identify areas for improvement. Big Team Science projects are often led by researchers from Anglo-Saxon and Global North institutions, while the contributions of researchers from the Global South are oftentimes diluted in the ordering of authors—i.e., authors from Global North tend to occupy positions of prestige such as the first, corresponding, and last author (e.g., refs. <sup>146,147,156–162</sup>) while researchers from Low- and Middle-income countries are compressed in the middle. Moreover, there are also challenges associated with collecting data in low-and-middle-income countries that are often not accounted for, such as limited access to polling infrastructure or technology and the gaping inequities in resources, funding, and educational opportunities. Furthermore, journals and research institutions do not always recognize contributions to Big Team Science projects, which can unequally negatively impact the academic careers of already marginalized researchers. For example, some prominent journals prefer mentioning consortium or group names instead of accommodating complete lists of author names in the byline, with the result that immediate author visibility decreases (see for example, ref. <sup>156</sup>). To promote social justice in Big Team Science practices, it is crucial to set norms that redistribute credit, resources, funds, and expertise rather than preserving the status quo of extractive intellectual labour. These complex issues must be examined carefully—and now—to ensure current customs do not unintendedly sustain colonialist, extractivist, and racist research practices that pervade academia and society at large. Big Team Science stakeholders must see to it that, as a minimum, Big Team Science doesn't perpetuate existing inequalities and power structures.

**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<sup>161,163–166</sup>, 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<sup>3,167</sup>. An example is the Transparent Psi Project<sup>168</sup> 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<sup>150</sup>. However, whether these initiatives contribute to research that is less biased in its central assumptions depends on wherein the "adversarial" nature lies. For example, two researchers may hold the same biased negative views towards one group, but opposing views on the implications of group membership on secondary outcomes. An adversarial collaboration from the same starting point pitching different methodological approaches and hypotheses about factors associated with group membership would still suffer from the same fundamental biases.

### 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 (d) diversity and inclusion as an area necessary to be considered in the context of open scholarship.

**Generalizability.** In extant work, the generalizability of effects is a serious concern (e.g., refs. <sup>169,170</sup>). 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 partially achieved 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<sup>170</sup>.

Possible solutions may lay in Big-team science and large-scale collaborations. 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<sup>149</sup>). 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., ref. <sup>171</sup>). 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<sup>149</sup>. At the same time, large collaborations might offer a chance to counteract geographical inequalities in research outputs<sup>172</sup>. 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, potentially increasing generalisability at various stages of the research process.

**Formal theory building.** Researchers have suggested that the replication crisis is, in fact, a "theory crisis"<sup>173</sup>. Low rates of replicability may be explained in part by the lack of formalism and first principles<sup>174</sup>. One example is the improper testing of theory or failures to identify auxiliary theoretical assumptions<sup>103,170</sup>. The verbal formulation of psychological theories and hypotheses cannot always be directly tested with

inferential statistics. Hence, generalisations provided in the literature are not always supported by the used data. Yarkoni<sup>170</sup> 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<sup>173</sup>. Indeed, many researchers may not even be at a stage where they are ready or able to test hypotheses<sup>175</sup>. Additional discussion of improving psychological theory and its evaluation is needed to advance the credibility revolution. Hence, a suggested approach to solving methodological problems is to (1) define variables, population parameters, and constants involved in the problem, including model assumptions, to then (2) formulate a formal mathematical problem statement. Results are (3) used to interrogate the problem. If the claims are valid, (4a) examples can be used to present practical relevance, while also (4b) presenting possible extensions and limitations. Finally, (5) policy making recommendations can be given. Additional discussion of improving 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<sup>174</sup>.

**Qualitative research.** Open scholarship research has focused primarily on quantitative data collection and analyses, with substantively less consideration for compatibility with qualitative or mixed methods<sup>51,176–178</sup>. 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<sup>179,180</sup>. 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; ref. <sup>41</sup>). 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<sup>181</sup>). 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<sup>6,179,182–184</sup>, and protocols are being developed specifically for qualitative research, such as preregistration templates<sup>185</sup>, practices increasing transparency<sup>186</sup>, and curating<sup>187</sup> and reusing qualitative data<sup>188</sup>. 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<sup>6,189</sup> is ongoing. Such an approach is reflected in various studies describing the tailored application of open scholarship protocols in qualitative studies<sup>182,184</sup>.

It is important to note that qualitative research also has dishes to add to the buffet of open science<sup>178,190</sup>. Qualitative research includes practices that realize forms of transparency that currently lack in quantitative work. One example is the practice of reflexivity, which aims to make transparent the positionality of the researcher(s) and their role in the production and interpretation of the data (see e.g., refs. <sup>191–193</sup>). A different form of transparency is 'member checking'<sup>194</sup>, which makes the participants in a study part of the analysis process by asking them to comment on a preliminary report of the analysis. These practices—e.g., member checking, positionality, reflexivity, critical team discussions, external audits and others—would likely



hold potential benefits for strictly quantitative research by promoting transparency and contextualization<sup>178,193</sup>.

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<sup>179,183</sup>. 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 the uptake of open scholarship practices.

**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<sup>51</sup>. For example, there are clear differences in the availability of resources by geographic region<sup>195,196</sup> and social groups, by ethnicity<sup>197</sup> or sex and gender<sup>51,198,199</sup>. 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<sup>195</sup>. Choosing (preferably free) open access options, including preprints and post-prints is one step allowing scholars to access resources irrespective of their privileges. An additional possibility is to waive article processing charges for researchers from low, or low-and-middle income countries. 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 embracing diversity<sup>150</sup>; 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 aiming to foster diversity and inclusion is essential.

### 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, if not impractical. While developments within the credibility revolution were originally fuelled by failed replications, these in themselves are not the only issue of discussion within the credibility revolution. Furthermore, replication rates alone may not be the best measure of research quality. Instead of focusing purely on replicability, we should strive to maximize transparency, rigour, and quality in all aspects of research<sup>18,200</sup>. To do so, we must observe structural, procedural, and community processes as intertwined drivers of change, and implement actionable changes on all levels. It is crucial that actors in different domains take responsibility for improvements and work together to ensure that high-quality outputs are incentivized and rewarded<sup>201</sup>. 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 meaningful reform will fail. In outlining multiple positive changes already implemented and embedded, we hope to provide our scientific community with hope, and a structure, to make further advances in the crises and revolutions to come.

Received: 20 January 2023; Accepted: 22 May 2023;

Published online: 25 July 2023

### References

- Bem, D. Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *J. Pers. Soc. Psychol.* **100**, 407 (2011).
- Crocker, J. The road to fraud starts with a single step. *Nature* **479**, 151–151 (2011).
- 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). *JPSP.* **100**, 426–432 (2011).
- Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 1–9 (2017).
- Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015). **This study was one of the first large-scale replication projects showing lower replication rates and smaller effect sizes among “successful” replicated findings.**
- Field, S. M., Hoekstra, R., Bringmann, L. & van Ravenzwaaij, D. When and why to replicate: as easy as 1, 2, 3? *Collabra Psychol.* **5**, 46 (2019).
- Nosek, B. A. et al. Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, 719–748 (2022). **This paper highlights the importance of addressing issues related to replicability, robustness, and reproducibility in psychological research to ensure the validity and reliability of findings.**
- Farrar, B. G., Boeckle, M. & Clayton, N. S. Replications in comparative cognition: what should we expect and how can we improve? *Anim. Behav. Cognit.* **7**, 1 (2020).
- Farrar, B. G., Voudouris, K. & Clayton, N. S. Replications, comparisons, sampling and the problem of representativeness in animal cognition research. *Anim. Behav. Cognit.* **8**, 273 (2021).
- Farrar, B. G. et al. Reporting and interpreting non-significant results in animal cognition research. *PeerJ* **11**, e14963 (2023).
- Errington, T. M. et al. Investigating the replicability of preclinical cancer biology. *Elife* **10**, e71601 (2021).
- Camerer, C. F. et al. Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).
- Frith, U. Fast lane to slow science. *Trends Cog. Sci.* **24**, 1–2 (2020).
- Pennington, C. A. *Student's Guide to Open Science: Using the Replication Crisis to Reform Psychology* (Open University Press, 2023).
- 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 Underst. Sci.* **29**, 270–288 (2020).
- Sanders, M., Snijders, V. & Hallsworth, M. Behavioural science and policy: where are we now and where are we going? *Behav. Public Policy* **2**, 144–167 (2018).
- Vazire, S. Implications of the credibility revolution for productivity, creativity, and progress. *Perspect. Psychol. Sci.* **13**, 411–417 (2018). **This paper explores how the rise of the credibility revolution, which emphasizes the importance of evidence-based knowledge and critical thinking, can lead to increased productivity, creativity, and progress in various fields.**
- Freese, J., Rauf, T. & Voelkel, J. G. Advances in transparency and reproducibility in the social sciences. *Soc. Sci. Res.* **107**, 102770 (2022).
- Trafimow, D. et al. Manipulating the alpha level cannot cure significance testing. *Front. Psychol.* **9**, 699 (2018).
- Loken, E. & Gelman, A. Measurement error and the replication crisis. *Science* **355**, 584–585 (2017).
- Azevedo, F. et al. Towards a culture of open scholarship: the role of pedagogical communities. *BMC Res. Notes* **15**, 75 (2022). **This paper details (a) the need to integrate open scholarship principles into research training within higher education; (b) the benefit of pedagogical communities and the role they play in fostering an inclusive culture of open scholarship; and (c) call for greater collaboration with pedagogical communities, paving the way for a much needed integration of top-down and grassroots open scholarship initiatives.**
- Grahe, J. E., Cuccolo, K., Leighton, D. C. & Cramblet Alvarez, L. D. Open science promotes diverse, just, and sustainable research and educational outcomes. *Psychol. Lean. Teach.* **19**, 5–20 (2020).
- Norris, E. & O'Connor, D. B. Science as behaviour: using a behaviour change approach to increase uptake of open science. *Psychol. Health* **34**, 1397–1406 (2019).
- Azevedo, F. et al. Introducing a framework for open and reproducible research training (FORRT). Preprint at <https://osf.io/bnh7p/> (2019). **This paper describes the importance of integrating open scholarship into higher education, its benefits and challenges, as well as about FORRT initiatives aiming to support educators in this endeavor.**
- Nuijten, M. B. & Polanin, J. R. “statcheck”: Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. *Res. Synth. Methods* **11**, 574–579 (2020).
- McAleer, P. et al. Embedding data skills in research methods education: preparing students for reproducible research. Preprint at <https://psyarxiv.com/hq68s/> (2022).

27. Holcombe, A. O., Kovacs, M., Aust, F. & Aczel, B. Documenting contributions to scholarly articles using CRediT and tenzing. *PLoS ONE* **15**, e0244611 (2020).
28. Koole, S. L. & Lakens, D. Rewarding replications: a sure and simple way to improve psychological science. *Perspect. Psychol. Sci.* **7**, 608–614 (2012).
29. Bauer, G. et al. Teaching constructive replications in the social sciences. Preprint at <https://osf.io/g3k5t/> (2022).
30. Wagge, J. R. et al. A demonstration of the Collaborative Replication and Education Project: Replication attempts of the red-romance effect. *Collabra Psychol.* **5**, 5 (2019). **A multi-institutional effort is being presented with the goal to replicate and teach research methods by collaboratively conducting and evaluating replications of three psychology experiments.**
31. Wagge, J. R. et al. Publishing research with undergraduate students via replication work: the collaborative replications and education project. *Front. Psychol.* **10**, 247 (2019).
32. Quintana, D. S. Replication studies for undergraduate theses to improve science and education. *Nat. Hum. Behav.* **5**, 1117–1118 (2021).
33. Button, K. S., Chambers, C. D., Lawrence, N. & Munafo, M. R. Grassroots training for reproducible science: a consortium-based approach to the empirical dissertation. *Psychol. Learn. Teach.* **19**, 77–90 (2020). **The article argues that improving the reliability and efficiency of scientific research requires a cultural shift in both thinking and practice, and better education in reproducible science should start at the grassroots, presenting a model of consortium-based student projects to train undergraduates in reproducible team science and reflecting on the pedagogical benefits of this approach.**
34. Feldman, G. Replications and extensions of classic findings in Judgment and Decision Making. <https://doi.org/10.17605/OSF.IO/5Z4A8> (2020). **A research team of early career researchers with the main activities in the years 2018–2023 focused on: 1) Mass scale project completing over 120 replications and extensions of classic findings in social psychology and judgment and decision making, 2) Building collaborative resources (tools, templates, and guides) to assist others in implementing open-science.**
35. Efendić, E. et al. Risky therefore not beneficial: replication and extension of Finucane et al.'s (2000) affect heuristic experiment. *Soc. Psychol. Personal. Sci.* **13**, 1173–1184 (2022).
36. Ziano, I., Yao, J. D., Gao, Y. & Feldman, G. Impact of ownership on liking and value: replications and extensions of three ownership effect experiments. *J. Exp. Soc. Psychol.* **89**, 103972 (2020).
37. Pownall, M. et al. Embedding open and reproducible science into teaching: a bank of lesson plans and resources. *Schol. Teach. Learn. Psychol.* (in-press) (2021). **To support open science training in higher education, FORRT compiled lesson plans and activities, and categorized them based on their theme, learning outcome, and method of delivery, which are made publicly available here: FORRT's Lesson Plans.**
38. Coles, N. A., DeBruine, L. M., Azevedo, F., Baumgartner, H. A. & Frank, M. C. 'big team' science challenges us to reconsider authorship. *Nat. Hum. Behav.* **7**, 665–667 (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. *Learn. Publ.* **32**, 71–74 (2019).
40. Allen, L., Scott, J., Brand, A., Hlava, M. & Altman, M. Publishing: Credit where credit is due. *Nature* **508**, 312–313 (2014).
41. Pownall, M. 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. *Roy. Soc. Open Sci.*, **10**, 221255 (2023). **This review article describes the available (empirical) evidence of the impact (and importance) of integrating open scholarship into higher education, its benefits and challenges on three specific areas: students' (a) scientific literacy; (b) engagement with science; and (c) attitudes towards science.**
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. *Teach. Psychol.* **45**, 158–163 (2018).
43. Frank, M. C. & Saxe, R. Teaching replication. *Perspect. Psychol. Sci.* **7**, 600–604 (2012). **In this perspective article, Frank and Saxe advocate for incorporating replication as a fundamental component of research training in psychology and other disciplines.**
44. Levin, N. & Leonelli, S. How does one "open" science? questions of value in biological research. *Sci. Technol. Human Values* **42**, 280–305 (2017).
45. Van Dijk, D., Manor, O. & Carey, L. B. Publication metrics and success on the academic job market. *Curr. Bio.* **24**, R516–R517 (2014).
46. Elsherif, M. M. et al. Bridging Neurodiversity and Open Scholarship: how shared values can Guide best practices for research integrity, social justice, and principled education. Preprint at <https://osf.io/preprints/metaarxiv/k7a9p/> (2022). **The authors describe systematic barriers, issues with disclosure, directions on prevalence and stigma, and the intersection of neurodiversity and open scholarship, and provide recommendations that can lead to personal and systematic changes to improve acceptance of neurodivergent individuals. Furthermore, perspectives of neurodivergent authors are being presented, the majority of whom have personal lived experiences of neurodivergence(s), and possible improvements in research integrity, inclusivity and diversity are being discussed.**
47. Onie, S. Redesign open science for Asia, Africa and Latin America. *Nature* **587**, 35–37 (2020).
48. Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D. & Mortenson, E. Racial inequality in psychological research: trends of the past and recommendations for the future. *Perspect. Psychol. Sci.* **15**, 1295–1309 (2020). **Roberts et al. highlight historical and current trends of racial inequality in psychological research and provide recommendations for addressing and reducing these disparities in the future.**
49. Steltenpohl, C. N. et al. Society for the improvement of psychological science global engagement task force report. *Collabra Psychol.* **7**, 22968 (2021).
50. Parsons, S. et al. A community-sourced glossary of open scholarship terms. *Nat. Hum. Behav.* **6**, 312–318 (2022). **In response to the varied and plural new terminology introduced by the open scholarship movement, which has transformed academia's lexicon, FORRT members have produced a community and consensus-based Glossary to facilitate education and effective communication between experts and newcomers.**
51. Pownall, M. et al. Navigating open science as early career feminist researchers. *Psychol. Women Q.* **45**, 526–539 (2021).
52. Gourdon-Kanhukamwe, A. et al. Opening up understanding of neurodiversity: a call for applying participatory and open scholarship practices. Preprint at <https://osf.io/preprints/metaarxiv/jq23s/> (2022).
53. Leech, G. Reversals in psychology. Behavioural and Social Sciences (Nature Portfolio) at <https://socialsciences.nature.com/posts/reversals-in-psychology> (2021).
54. Orben, A. A journal club to fix science. *Nature* **573**, 465–466 (2019).
55. Arnold, B. et al. The turing way: a handbook for reproducible data science. *Zenodo* <https://doi.org/10.5281/zenodo.3233986> (2019).
56. Open Life Science. A mentoring & training program for Open Science ambassadors. <https://openlifesci.org/> (2023).
57. Almarzouq, B. et al. Opensciency—a core open science curriculum by and for the research community (2023).
58. Schönbrodt, F. et al. Netzwerk der Open-Science-Initiativen (NOSI). <https://osf.io/tbkzh/> (2016).
59. Ball, R. et al. Course Syllabi for Open and Reproducible Methods. <https://osf.io/vkhhb/> (2022).
60. The Carpentries. <https://carpentries.org/> (2023).
61. The Embassy of Good Science. [https://embassy.science/wiki/Main\\_Page](https://embassy.science/wiki/Main_Page) (2023).
62. Berkeley Initiative for Transparency in the Social Sciences. <https://www.bitsc.org/> (2023).
63. Institute for Replication. <https://i4replication.org/> (2023).
64. Reproducibility for Everyone. <https://www.repro4everyone.org/> (2023).
65. Armeni, K. et al. Towards wide-scale adoption of open science practices: The role of open science communities. *Sci. Public Policy* **48**, 605–611 (2021).
66. Welcome to the UK Reproducibility Network The UK Reproducibility Network (UKRN). <https://www.ukrn.org/> (2023).
67. Collyer, F. M. Global patterns in the publishing of academic knowledge: Global North, global South. *Curr. Soc.* **66**, 56–73 (2018).
68. Ali-Khan, S. E., Harris, L. W. & Gold, E. R. Motivating participation in open science by examining researcher incentives. *Elife* **6**, e29319 (2017).
69. Robson, S. G. et al. Promoting open science: a holistic approach to changing behaviour. *Collabra Psychol.* **7**, 30137 (2021).
70. Coalition for Advancing Research Assessment. <https://coara.eu/> (2023).
71. Vanclay, J. K. Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics* **92**, 211–238 (2012).
72. Kidwell, M. C. et al. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Bio.* **14**, e1002456 (2016).
73. Rowhani-Farid, A., Aldcroft, A. & Barnett, A. G. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Roy. Soc. Open Sci.* **7**, 191818 (2020).
74. Thibault, R. T., Pennington, C. R. & Munafo, M. Reflections on preregistration: core criteria, badges, complementary workflows. *J. Trial & Err.* <https://doi.org/10.36850/mr6> (2022).
75. Chambers, C. D. Registered reports: a new publishing initiative at Cortex. *Cortex* **49**, 609–610 (2013).
76. Chambers, C. D. & Tzavella, L. The past, present and future of registered reports. *Nat. Hum. Behav.* **6**, 29–42 (2022).
77. Soderberg, C. K. et al. Initial evidence of research quality of registered reports compared with the standard publishing model. *Nat. Hum. Behav.* **5**, 990–997 (2021).
78. Scheel, A. M., Schijen, M. R. & Lakens, D. An excess of positive results: comparing the standard Psychology literature with Registered Reports. *Adv. Meth. Pract. Psychol. Sci.* **4**, 25152459211007467 (2021).

79. Renbarger, R. et al. Champions of transparency in education: what journal reviewers can do to encourage open science practices. Preprint at <https://doi.org/10.35542/osf.io/xqfwb>.
80. Nosek, B. A. et al. Transparency and Openness Promotion (TOP) Guidelines. Center for Open Science project. <https://osf.io/9f6gx/> (2022).
81. Silverstein, P. et al. A Guide for Social Science Journal Editors on Easing into Open Science. (2023). Preprint at <https://doi.org/10.31219/osf.io/hstcx>.
82. NASA. Transform to Open Science (TOPS). <https://github.com/nasa/Transform-to-Open-Science> (2023).
83. UNESCO. UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en> (2021).
84. European University Association. <https://eua.eu> (2023).
85. Munafo, M. R. Improving the efficiency of grant and journal peer review: registered reports funding. *Nicotine Tob. Res.* **19**, 773–773 (2017).
86. Else, H. A guide to Plan S: the open-access initiative shaking up science publishing. *Nature* (2021).
87. Mills, M. Plan S—what is its meaning for open access journals and for the JACMP? *J. Appl. Clin. Med. Phys.* **20**, 4 (2019).
88. Zhang, L., Wei, Y., Huang, Y. & Sivertsen, G. Should open access lead to closed research? the trends towards paying to perform research. *Scientometrics* **127**, 7653–7679 (2022).
89. McNutt, M. Plan S falls short for society publishers—and for the researchers they serve. *Proc. Natl Acad. Sci. USA* **116**, 2400–2403 (2019).
90. PeerCommunityIn. <https://peercommunityin.org/> (2023).
91. Elife. <https://elifesciences.org/for-the-press/b2329859/elife-ends-accept-reject-decisions-following-peer-review> (2023).
92. Nosek, B. A., Spies, J. R. & Motyl, M. Scientific utopia II: Restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* **7**, 615–631 (2012).
93. Schönbrodt, F. <https://www.nicebread.de/open-science-hiring-practices/> (2016).
94. Delios, A. et al. Examining the generalizability of research findings from archival data. *Proc. Natl Acad. Sci. USA* **119**, e2120377119 (2022).
95. Dreber, A. et al. Using prediction markets to estimate the reproducibility of scientific research. *Proc. Natl Acad. Sci. USA* **112**, 15343–15347 (2015).
96. Fraser, H. et al. Predicting reliability through structured expert elicitation with the replicATS (Collaborative Assessments for Trustworthy Science) process. *PLoS ONE* **18**, e0274429 (2023).
97. 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**, e0248780 (2021).
98. Tierney, W. et al. Creative destruction in science. *Organ. Behav. Hum. Decis. Process* **161**, 291–309 (2020).
99. Tierney, W. et al. A creative destruction approach to replication: implicit work and sex morality across cultures. *J. Exp. Soc. Psychol.* **93**, 104060 (2021).
100. Hoogeveen, S., Sarafoglou, A. & Wagenmakers, E.-J. Laypeople can predict which social-science studies will be replicated successfully. *Adv. Meth. Pract. Psychol. Sci.* **3**, 267–285 (2020).
101. Lewandowsky, S. & Oberauer, K. Low replicability can support robust and efficient science. *Nat. Commun.* **11**, 358 (2020).
102. Button, K. S. & Munafo, M. R. in *Psychological Science under Scrutiny: Recent Challenges and Proposed Solutions* 22–33 (2017).
103. Świątkowski, W. & Dompnier, B. Replicability crisis in social psychology: looking at the past to find new pathways for the future. *Int. Rev. Soc. Psychol.* **30**, 111–124 (2017).
104. Simonsohn, U., Nelson, L. D. & Simmons, J. P. P-curve: a key to the file-drawer. *J. Exp. Psychol. Gen.* **143**, 534 (2014).
105. Brunner, J. & Schimmack, U. Estimating population mean power under conditions of heterogeneity and selection for significance. *Meta-Psychol.* **4**, 1–22 (2020).
106. Benjamin, D. J. et al. Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2018).
107. Rubin, M. & Donkin, C. Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philos. Psychol.* (in-press) 1–29 (2022).
108. Amrhein, V. & Greenland, S. Remove, rather than redefine, statistical significance. *Nat. Hum. Behav.* **2**, 4–4 (2018).
109. Trafimow, D. & Marks, M. Editorial in basic and applied social psychology. *Basic Appl. Soc. Psych.* **37**, 1–2 (2015).
110. Lakens, D. et al. Justify your alpha. *Nat. Hum. Behav.* **2**, 168–171 (2018).
111. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological research: a tutorial. *Adv. Meth. Pract. Psychol. Sci.* **1**, 259–269 (2018).
112. Verhagen, J. & Wagenmakers, E.-J. Bayesian tests to quantify the result of a replication attempt. *J. Exp. Psychol. Gen.* **143**, 1457 (2014).
113. Dienes, Z. Bayesian versus orthodox statistics: Which side are you on?. *Perspect. Psychol. Sci.* **6**, 274–290 (2011).
114. Love, J. et al. JASP: Graphical statistical software for common statistical designs. *J. Stat. Soft.* **88**, 1–17 (2019).
115. Şahin, M. & Aybek, E. Jamovi: an easy to use statistical software for the social scientists. *Int. J. Assess. Tools Educ.* **6**, 670–692 (2019).
116. Brown, N. J. & Heathers, J. A. The GRIM test: a simple technique detects numerous anomalies in the reporting of results in psychology. *Soc. Psychol. Personal. Sci.* **8**, 363–369 (2017).
117. Heathers, J. A., Anaya, J., van der Zee, T. & Brown, N. J. Recovering data from summary statistics: Sample parameter reconstruction via iterative techniques (SPRITE). *PeerJ Preprints* **6**, e26968v1 (2018).
118. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).
119. Breznau, N. et al. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proc. Natl Acad. Sci. USA* **119**, e2203150119 (2022).
120. Breznau, N. et al. The Crowdsourced Replication Initiative: investigating immigration and social policy preferences: executive report. <https://osf.io/preprints/socarxiv/6j9qb/> (2019).
121. 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).
122. Azevedo, F. & Jost, J. T. The ideological basis of antiscientific attitudes: effects of authoritarianism, conservatism, religiosity, social dominance, and system justification. *Group Process. Intergr. Relat.* **24**, 518–549 (2021).
123. Heininga, V. E., Oldehinkel, A. J., Veenstra, R. & Nederhof, E. I just ran a thousand analyses: benefits of multiple testing in understanding equivocal evidence on gene-environment interactions. *PLoS ONE* **10**, e0125383 (2015).
124. Liu, Y., Kale, A., Althoff, T. & Heer, J. Boba: Authoring and visualizing multiverse analyses. *IEEE Trans. Vis. Comp. Graph.* **27**, 1753–1763 (2020).
125. Harder, J. A. The multiverse of methods: extending the multiverse analysis to address data-collection decisions. *Perspect. Psychol. Sci.* **15**, 1158–1177 (2020).
126. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).
127. Azevedo, F., Marques, T. & Micheli, L. In pursuit of racial equality: identifying the determinants of support for the black lives matter movement with a systematic review and multiple meta-analyses. *Perspect. Politics*, (in-press), 1–23 (2022).
128. Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. *Introduction to Meta-analysis* (John Wiley & Sons, 2021).
129. Higgins, J. P. et al. *Cochrane Handbook for Systematic Reviews of Interventions* (John Wiley & Sons, 2022).
130. Carter, E. C., Schönbrodt, F. D., Gervais, W. M. & Hilgard, J. Correcting for bias in psychology: a comparison of meta-analytic methods. *Adv. Meth. Pract. Psychol. Sci.* **2**, 115–144 (2019).
131. 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). *Behav. Res. Methods* **48**, 1205–1226 (2016).
132. Van Assen, M. A., van Aert, R. & Wicherts, J. M. Meta-analysis using effect size distributions of only statistically significant studies. *Psychol. Meth.* **20**, 293 (2015).
133. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int. J. Surg.* **88**, 105906 (2021).
134. Topor, M. K. et al. An integrative framework for planning and conducting non-intervention, reproducible, and open systematic reviews (NIRO-SR). *Meta-Psychol.* (In Press) (2022).
135. Van den Akker, O. et al. Generalized systematic review registration form. Preprint at <https://doi.org/10.31222/osf.io/3nbea> (2020).
136. Booth, A. et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Sys. Rev.* **1**, 1–9 (2012).
137. Cristea, I. A., Naudet, F. & Caquelin, L. Meta-research studies should improve and evaluate their own data sharing practices. *J. Clin. Epidemiol.* **149**, 183–189 (2022).
138. Knobloch, K., Yoon, U. & Vogt, P. M. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J. Craniomaxillofac Surg.* **39**, 91–92 (2011).
139. Lakens, D., Hilgard, J. & Staaks, J. On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychol.* **4**, 1–10 (2016).
140. Editors, P. M. Best practice in systematic reviews: the importance of protocols and registration. *PLoS Med.* **8**, e1001009 (2011).
141. Tsujimoto, Y. et al. Majority of systematic reviews published in high-impact journals neglected to register the protocols: a meta-epidemiological study. *J. Clin. Epidemiol.* **84**, 54–60 (2017).
142. Xu, C. et al. Protocol registration or development may benefit the design, conduct and reporting of dose-response meta-analysis: empirical evidence from a literature survey. *BMC Med. Res. Meth.* **19**, 1–10 (2019).
143. Polanin, J. R., Hennessy, E. A. & Tsuji, S. Transparency and reproducibility of meta-analyses in psychology: a meta-review. *Perspect. Psychol. Sci.* **15**, 1026–1041 (2020).

144. Uhlmann, E. L. et al. Scientific utopia III: Crowdsourcing science. *Perspect. Psychol. Sci.* **14**, 711–733 (2019).
145. So, T. Classroom experiments as a replication device. *J. Behav. Exp. Econ.* **86**, 101525 (2020).
146. Ebersole, C. R. et al. Many labs 3: evaluating participant pool quality across the academic semester via replication. *J. Exp. Soc. Psychol.* **67**, 68–82 (2016).
147. Klein, R. et al. Investigating variation in replicability: a “many labs” replication project. *Soc. Psychol.* **45**, 142–152 (2014).
148. Glöckner, A. et al. Hagen Cumulative Science Project. Project overview at [osf.io/d7za8](https://osf.io/d7za8) (2015).
149. Moshontz, H. et al. The psychological science accelerator: advancing psychology through a distributed collaborative network. *Adv. Meth. Pract. Psychol. Sci.* **1**, 501–515 (2018).
150. Forscher, P. S. et al. The Benefits, Barriers, and Risks of Big-Team Science. *Perspect. Psychol. Sci.* **17**, 17456916221082970 (2020). **The paper discusses the advantages and challenges of conducting large-scale collaborative research projects, highlighting the potential for increased innovation and impact, as well as the difficulties in managing complex collaborations and addressing issues related to authorship and credit.**
151. Lieck, D. S. N. & Lakens, D. An Overview of Team Science Projects in the Social Behavioral Sciences. <https://doi.org/10.17605/OSF.IO/WX4ZD> (2022).
152. Jarke, H. et al. A roadmap to large-scale multi-country replications in psychology. *Collabra Psychol.* **8**, 57538 (2022).
153. 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. *Exp. Clin. Psychopharmacol.* **30**, 444–451 (2022).
154. Disis, M. L. & Slattery, J. T. The road we must take: multidisciplinary team science. *Sci. Trans. Med.* **2**, 22cm9–22cm9 (2010).
155. Ledgerwood, A. et al. The pandemic as a portal: reimagining psychological science as truly open and inclusive. *Perspect. Psychol. Sci.* **17**, 937–959 (2022).
156. Legate, N. et al. A global experiment on motivating social distancing during the COVID-19 pandemic. *Proc. Natl Acad. Sci. USA* **119**, e2111091119 (2022).
157. Nexus, P. N. A. S. Predicting attitudinal and behavioral responses to COVID-19 pandemic using machine learning. *Proc. Natl Acad. Sci. USA* **1**, 1–15 (2022).
158. Van Bavel, J. J. et al. National identity predicts public health support during a global pandemic. *Nat. Commun.* **13**, 517 (2022).
159. Buchanan, E. M. et al. The psychological science accelerator’s COVID-19 rapid-response dataset. *Sci. Data* **10**, 87 (2023).
160. Azevedo, F. et al. Social and moral psychology of covid-19 across 69 countries. *Nat. Sci. Dat.* <https://kar.kent.ac.uk/99184/> (2022).
161. Wang, K. et al. A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. *Nat. Hum. Behav.* **5**, 1089–1110 (2021).
162. Dorison, C. A. et al. In COVID-19 health messaging, loss framing increases anxiety with little-to-no concomitant benefits: Experimental evidence from 84 countries. *Affect. Sci.* **3**, 577–602 (2022).
163. Coles, N. A. et al. A multi-lab test of the facial feedback hypothesis by the many smiles collaboration. *Nat. Hum. Behav.* **6**, 1731–1742 (2022).
164. 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. *J. Pers. Soc. Psychol.* **124**, 287 (2023).
165. Cowan, N. et al. How do scientific views change? notes from an extended adversarial collaboration. *Perspect. Psychol. Sci.* **15**, 1011–1025 (2020).
166. Forscher, P. S. et al. Stereotype threat in black college students across many operationalizations. Preprint at <https://psyarxiv.com/6hju9/> (2019).
167. Kahneman, D. & Klein, G. Conditions for intuitive expertise: a failure to disagree. *Am. Psychol.* **64**, 515 (2009).
168. Kekecs, Z. et al. Raising the value of research studies in psychological science by increasing the credibility of research reports: the transparent Psi project. *Roy. Soc. Open Sci.* **10**, 191375 (2023).
169. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
170. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* **45**, e1 (2022).
171. Ghai, S. It’s time to reimagine sample diversity and retire the WEIRD dichotomy. *Nat. Hum. Behav.* **5**, 971–972 (2021). **The paper argues that the reliance on WEIRD (Western, educated, industrialized, rich, and democratic) samples in psychological research limits the generalizability of findings and suggests reimagining sample diversity to ensure greater external validity.**
172. Nielsen, M. W. & Andersen, J. P. Global citation inequality is on the rise. *Proc. Natl Acad. Sci. USA* **118**, e2012208118 (2021).
173. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* **26**, 1596–1618 (2019).
174. Devezer, B., Navarro, D. J., Vandekerckhove, J. & Ozge Buzbas, E. The case for formal methodology in scientific reform. *Roy. Soc. Open Sci.* **8**, 200805 (2021).
175. Scheel, A. M., Tiokhin, L., Isager, P. M. & Lakens, D. Why hypothesis testers should spend less time testing hypotheses. *Perspect. Psychol. Sci.* **16**, 744–755 (2021).
176. Chauvette, A., Schick-Makaroff, K. & Molzahn, A. E. Open data in qualitative research. *Int. J. Qual. Meth.* **18**, 1609406918823863 (2019).
177. Field, S. M., van Ravenzwaaij, D., Pittelkow, M.-M., Hoek, J. M. & Derksen, M. Qualitative open science—pain points and perspectives. Preprint at <https://doi.org/10.31219/osf.io/e3cq4> (2021).
178. Steltenpohl, C. N. et al. Rethinking transparency and rigor from a qualitative open science perspective. *J. Trial & Err.* <https://doi.org/10.36850/mr7> (2023).
179. Branney, P. et al. Three steps to open science for qualitative research in psychology. *Soc. Pers. Psy. Comp.* **17**, 1–16 (2023).
180. VandeVusse, A., Mueller, J. & Karcher, S. Qualitative data sharing: Participant understanding, motivation, and consent. *Qual. Health Res.* **32**, 182–191 (2022).
181. Çelik, H., Baykal, N. B. & Memur, H. N. K. Qualitative data analysis and fundamental principles. *J. Qual. Res. Educ.* **8**, 379–406 (2020).
182. Class, B., de Bruyne, M., Wullemin, C., Donzé, D. & Claivaz, J.-B. Towards open science for the qualitative researcher: from a positivist to an open interpretation. *Int. J. Qual. Meth.* **20**, 16094069211034641 (2021).
183. Humphreys, L., Lewis Jr, N. A., Sender, K. & Won, A. S. Integrating qualitative methods and open science: five principles for more trustworthy research. *J. Commun.* **71**, 855–874 (2021).
184. Steinhardt, I., Bauer, M., Wünsche, H. & Schimmler, S. The connection of open science practices and the methodological approach of researchers. *Qual. Quant.* (in-press) 1–16 (2022).
185. Haven, T. L. & Van Grootel, L. Preregistering qualitative research. *Account. Res.* **26**, 229–244 (2019).
186. Frohwrth, L., Karcher, S. & Lever, T. A. A transparency checklist for qualitative research. Preprint at <https://doi.org/10.31235/osf.io/wc35g> (2023).
187. Demgenski, R., Karcher, S., Kirilova, D. & Weber, N. Introducing the qualitative data repository’s curation handbook. *J. eSci. Librariansh.* **10**, 1–11 (2021).
188. Karcher, S., Kirilova, D., Pagé, C. & Weber, N. How data curation enables epistemically responsible reuse of qualitative data. *Qual. Rep.* **26**, 1996–2010 (2021).
189. Bergmann, C. How to integrate open science into language acquisition research? Student workshop at BUCLD 43 (2018).
190. Bergmann, C. The buffet approach to open science. <https://cogtales.wordpress.com/2023/04/16/the-buffet-approach-to-open-science/> (2023).
191. Field, S. M. & Derksen, M. Experimenter as automaton; experimenter as human: exploring the position of the researcher in scientific research. *Eur. J. Philos. Sci.* **11**, 11 (2021).
192. Chenail, R. J. Communicating your qualitative research better. *Fam. Bus. Rev.* **22**, 105–108 (2009).
193. Levitt, H. M. et al. The meaning of scientific objectivity and subjectivity: from the perspective of methodologists. *Psychol. Methods* **27**, 589–605 (2020).
194. Candela, A. G. Exploring the function of member checking. *Qual. Rep.* **24**, 619–628 (2019).
195. Petersen, O. H. Inequality of research funding between different countries and regions is a serious problem for global science. *Function* **2**, zqab060 (2021).
196. Puthillam, A. et al. Guidelines to improve internationalization in psychological science. Preprint at <https://psyarxiv.com/2u4h5/> (2022).
197. Taffe, M. & Gilpin, N. Equity, diversity and inclusion: racial inequity in grant funding from the US National Institutes of Health. *eLife* **10**, e65697 (2021).
198. 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 Med.* **16**, e1002935 (2019).
199. Sato, S., Gygas, P. M., Randall, J. & Schmid Mast, M. The leaky pipeline in research grant peer review and funding decisions: challenges and future directions. *High. Educ.* **82**, 145–162 (2021).
200. Guttinger, S. The limits of replicability. *Eur. J. Philos. Sci.* **10**, 10 (2020).
201. Evans, T. Developments in open data norms. *J. Open Psychol. Data* **10**, 1–6 (2022).
202. John, L. K., Loewenstein, G. & Prelec, D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* **23**, 524–532 (2012).
203. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2016).
204. Wicherts, J. M. et al. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. *Front. Psychol.* **7**, 1832 (2016).
205. Flake, J. K., Pek, J. & Hehman, E. Construct validation in social and personality research: current practice and recommendations. *Soc. Psychol. Personal. Sci.* **8**, 370–378 (2017).

206. Flake, J. K. & Fried, E. I. Measurement schmeasurement: questionable measurement practices and how to avoid them. *Adv. Meth. Pract. Psychol. Sci.* **3**, 456–465 (2020).
207. Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P. & Cubelli, R. Questionable research practices among Italian research psychologists. *PLoS ONE* **12**, e0172792 (2017).
208. Fiedler, K. & Schwarz, N. Questionable research practices revisited. *Soc. Psychol. Personal. Sci.* **7**, 45–52 (2016).
209. Kerr, N. L. HARKing: Hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.* **2**, 196–217 (1998).
210. 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* 744–751 (Springer, 2022).
211. Gerrits, R. G. et al. 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**, e027903 (2019).
212. Checchi, D., De Fraja, G. & Verzillo, S. Incentives and careers in academia: theory and empirical analysis. *Rev. Econ. Stat.* **103**, 786–802 (2021).
213. Grove, L. *The Effects of Funding Policies on Academic Research*. Ph.D. thesis, University College London (2017).
214. Frias-Navarro, D., Pascual-Soler, M., Perezgonzalez, J., Monderde-i Bort, H. & Pascual-Llobell, J. Spanish Scientists' Opinion about Science and Researcher Behavior. *Span. J. Psychol.* **24**, e7 (2021).
215. Bornmann, L. & Daniel, H.-D. The state of h index research: is the h index the ideal way to measure research performance? *EMBO Rep.* **10**, 2–6 (2009).
216. Munafo, M. et al. Scientific rigor and the art of motorcycle maintenance. *Nat. Biotechnol.* **32**, 871–873 (2014).
217. Primbs, M. A. et al. Are small effects the indispensable foundation for a cumulative psychological science? A reply to Götz et al. (2022). *Perspect. Psychol. Sci.* **18**, 508–512 (2022).
218. Martin, G. & Clarke, R. M. Are psychology journals anti-replication? A snapshot of editorial practices. *Front. Psychol.* **8**, 523 (2017).
219. Cohen, B. A. How should novelty be valued in science? *Elife* **6**, e28699 (2017).
220. Tjldink, J. K., Vergouwen, A. C. & Smulders, Y. M. Publication pressure and burn out among Dutch medical professors: a nationwide survey. *PLoS ONE* **8**, e73381 (2013).
221. Tjldink, J. K., Verbeke, R. & Smulders, Y. M. Publication pressure and scientific misconduct in medical scientists. *J. Empir. Res. Hum. Res. Ethics* **9**, 64–71 (2014).
222. Laitin, D. D. et al. Reporting all results efficiently: a RARE proposal to open up the file drawer. *Proc. Natl Acad. Sci. USA* **118**, e2106178118 (2021).
223. Franco, A., Malhotra, N. & Simonovits, G. Publication bias in the social sciences: unlocking the file drawer. *Science* **345**, 1502–1505 (2014).
224. Matarese, V. Kinds of replicability: different terms and different functions. *Axiomathes* 1–24 (2022).
225. Maxwell, S. E., Lau, M. Y. & Howard, G. S. Is psychology suffering from a replication crisis? what does “failure to replicate” really mean? *Am. Psychol.* **70**, 487 (2015).
226. Ulrich, R. & Miller, J. Questionable research practices may have little effect on replicability. *Elife* **9**, e58237 (2020).
227. Devezer, B. & Buzbas, E. Minimum viable experiment to replicate (2021). Preprint at <http://philsci-archive.pitt.edu/21475/>.
228. Stroebe, W. & Strack, F. The alleged crisis and the illusion of exact replication. *Perspect. Psychol. Sci.* **9**, 59–71 (2014).
229. Feest, U. Why replication is overrated. *Phil. Sci.* **86**, 895–905 (2019).
230. Eronen, M. I. & Bringmann, L. F. The theory crisis in psychology: how to move forward. *Perspect. Psychol. Sci.* **16**, 779–788 (2021).

## Acknowledgements

We want to thank Ali H. Al-Hoorie for his contribution with reference formatting and commenting on the manuscript, and Sriraj Aiyer, Crystal Steltenpohl, Maarten Derksen,

Rinske Vermeij, and Esther Plomp for the comments provided on an earlier version of the manuscript. No funding has been received for this work.

## Author contributions

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<sup>27</sup>. Conceptualization: M.K., F.A. and T.R.E. Methodology: M.K., F.A. and T.R.E. Project administration: M.K. and F.A. Visualisation: M.K. and F.A. Writing—original draft: M.K., F.A., T.R.E., C.R.P. Writing—review and editing: M.K., F.A., H.H., M.M.E., N.B., A.O.M., T.K., J.Ø.S.O., C.R.P., J.J.S., M.P., B.G., Y.Y., J.P.R., J.M., S.A., B.J.B., S.G., C.M.O., T.W., S.K.Y., M.L., L.M.K., N.A.A., O.L., T.R.E., O.R., L.M., K.S. and S.Y.

## Competing interests

This work is as an initiative from The Framework for Open and Reproducible Research Training (FORRT; <https://forrt.org>). All authors are advocates of open scholarship with different open science organisational affiliations, including FORRT: M.K., F.A., H.H., M.M.E., A.O.M., B.S., T.K., C.R.P., J.J.S., Y.Y., J.P.R., J.M., S.A., B.J.B., S.G., C.M.O., T.W., S.K.Y., M.L., L.M.K., N.A.A., O.L., T.E., O.R., L.M., K.S. and S.Y. ReproducibiliTEA journal clubs: Y.Y., J.S., C.R.P., M.M.E. and O.L. NASA TOPs: F.A. and S.A. Student Initiative for Open Science: M.K. Open Applied Linguistics: M.L. Psychological Science Accelerator: M.K., F.A., T.K., C.R.P., M.P., K.S., M.E., N.B., B.B., A.O.M., B.G., Y.Y., J.P.R., S.A., S.K.Y., M.L., N.A.A., T.E. and M.M.E. ManyMany: B.G. and M.M.E. ManyBabies: M.M.E. The United Kingdom Reproducibility Network: F.A., C.R.P., T.E., O.R., C.R.P. and A.O.M. Reproducible Research Oxford: O.R. The Sports Science Replication Centre: J.M. Norway's Reproducibility Network: T.K. and M.K. Interessengruppe Offene und Reproduzierbare Forschung: H.H. Arbeitsgruppe Open Science der Fachgruppe Gesundheitspsychologie in der Deutschen Gesellschaft für Psychologie: L.M.K. Reproducible Interpretable Open Transparent (RIOT) Science Club: T.K. Center for Open Science: S.A. Open Science: F.A. and S.A. Society for the Improvement of Psychological Science: F.A. and S.A. Stanford Center for Open and Reproducible Science: S.A.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44271-023-00003-2>.

**Correspondence** and requests for materials should be addressed to Flavio Azevedo.

**Peer review information** *Communications Psychology* thanks Rima-Maria Rahal and Ulf Toelch for their contribution to the peer review of this work. Primary Handling Editor: Marike Schiffer. A peer review file is available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023