How open science norms improve scientific practices

Ljiljana B. Lazarević[1], Iris Žeželj[2]
[1]*Institute of psychology and Laboratory for research of Individual Differences, University of Belgrade, Belgrade, Serbia*
[2]*Department of psychology, University of Belgrade, Belgrade, Serbia*
ljiljana.lazarevic@f.bg.ac.rs, izezelj@f.bg.ac.rs

Abstract: Empirical findings indicate that various scientific fields are in serious replication crisis. One source of low reproducibility can be attributed to the so-called Questionable Research Practices. The wider incentive structures prioritizing novel and positive findings over negative ones also influence the rate of false-positive effects in published papers. We are discussing the open science norms and good scientific practices targeted to improve methods, reporting, dissemination and incentives for researchers. Among proposed norms are pre-registration of study protocols, results-blind peer reviews, collaboration and team-science, and improvement of statistical and methodological training. Moreover, important stakeholders are recognizing the importance of open science, and incentives for researchers to openly share work are developing. Finally, we are discussing some of the remaining challenges in the quest for open, reproducible, and credible science.

Keywords: open science; reproducibility; questionable scientific practices; good scientific practices.

I. Introduction

Ideally, science should be an unbiased, collaborative quest for truth, for the benefit of the whole society. To accomplish that ideal, it should be cumulative and self-corrective. It means that over time, credible scientific findings should be independently replicated and preserved, while so-called false-positive findings should be refuted [1]. It is important to note that the self-correction principle does not imply necessarily that all published science is credible and replicable – it implies that it only such findings should stand the test of time. The estimate of the proportion of the correct first-proposed results varies between 0% and 100%. However, more importantly, the proportion of the correct results after meta-analytically analyzing *all available evidence* still varies between 0% and 100% [2].

Novel analyses from different empirical scientific fields, revealed a serious replication crisis [3][4]. In psychology, a large-scale collaboration successfully replicated only around 50% of the selected published studies [4][5]. A recent high-powered effort replicating experimental studies in the social sciences published in the most prestigious outlets - Science and Nature - between 2010 and 2015 reported that the average effect size of the replications was 50% of the original effect size [6]. In economics, 39% of the replicated studies did not have significant effects in the same direction as original studies [7]. Having public health in mind, the most disturbing estimate comes from the biomedical research: the percentage of non-reproducible findings ranges from 75% to 90% [8].

It seemed that the climate in empirical science did not foster cumulative, reproducible knowledge. There were no mechanisms in place that would guarantee the necessary self-correctiveness. In fact, it was just the oposite.

II. Questionable Research Practices

Scientific practices that contributed to such sad state of affairs were somewhat euphemistically labelled Questionable Research Practices (QRP) [2][9]. Studies show that low reproducibility rate of scientific findings resulted from a manifold of practices: too small sample sizes, searching for statistically significant results (p-hacking), post hoc hypothesizing (HARK-ing: Hypothesizing After the Results are Known), diverging from originally planned research designs and analytical strategies, selectively reporting studies with positive results and not studies with negative results, etc. [2][3][10][11].

It would not be fair to attribute the responsibility for such practices solely to scientists: it was the system that rewarded flashy positive findings and marginalized negative ones. That resulted in self-censoring by scientists, further biased selection by journal editors, project funders and employers, which inevitably lead to a grossly inflated rate of false-positive effects in published papers [11][12].

If the system is to blame, then the system needs to change, and there seemed to be a substantial room for improvement in research and publishing practices. So far, a number of innovations were introduced to this end (for an overview, see [11]).

III. Open science norms and good scientific practices

The credibility of scientific findings depends on the transparency of all elements of the scientific process [13]. In this paper, we will briefly present the most prominent practices aimed to improve methods, reporting, dissemination and incentives for research.

A. Methods

As science is a human endeavour, scientists are prone to biases in formulating hypotheses, conducting data collection and data preparation - they typically, and most often not intentionally, favour confirming their starting hypotheses [11][14]. As a countermeasure, a *pre-registration* of the study design, hypothesis, analytic strategy, and measured outcomes to the journal for peer review before beginning the study was proposed [11][15]. Some scientific disciplines, like clinical medicine or stomatology, introduced a mandatory pre-registration of study protocols to address publication bias and analytical flexibility [11][16][17]. In social and behavioural sciences, pre-registration is becoming more frequent: the journals are adopting and promoting the Registered Reports publishing format encouraging pre-registration and results-blind peer review [11].

In addition, in an attempt to avoid publication biases, journals are now introducing a policy of *results blind evaluation* of manuscripts submitted to journals [18], for the illustration of the process, see [19][20]. Results blind evaluation implies that the review process is two-staged: in the first step, the editor distributes the manuscript containing the Introduction and the Methods section for external review; if the decision of the first stage review is positive, it proceeds to the second step, in which the full manuscript is once again evaluated.

To address the problem of low statistical power (too small samples to allow reliable conclusions) which increases the likelihood of obtaining both false-positive and false-negative results [21], researchers suggested to join forces. C*ollaboration* across many sites enabled high-powered study designs, stringent analytic strategy, and more economic use of resources. A very good example of efficient collaborations in social and behavioural sciences are The Many Labs projects [5][22][23] or other large-scale collaborations (e.g., [24]); Psychological Science Accelerator: [25]. In other disciplines, there are older and more prominent examples: The LIGO scientific collaboration [26]; The GENOME project [27]; The CERN project [28].

Finally, most "reformers" agree that *more rigorous statistical and methodological training* and research methods is fundamental for high-quality research. Although most of the research programs do contain these courses, they should be fundamentally reinvented; some authors argue that "departments need to begin teaching statistical thinking, not rituals" ([29], p. 198). As a good example of a crowdsourced project with an emphasis on the replication and pedagogy we can list an initiative, launched in 2012, *Collaborative Replications and Education Project* (CREP; https://osf.io/wfc6u/) aimed to strengthen undergraduate students' knowledge and expertise in statistics and research methods.

B. Reporting and Dissemination

First global initiatives for more sharing and transparency appeared almost 20 years ago. In 2002, the Budapest Open Access Initiative (BOAI) gathered a diverse group of stakeholders and launched a worldwide campaign for open access (OA) to all peer-reviewed research. A year later, the Berlin Declaration on open access, signed by nearly 300 research institutions, funding bodies, libraries, museums, and governments, was based on the BOAI and called for the research results to be publicly available.

In recent years, important stakeholders started demanding from *researchers to share materials, databases and analytical scripts*. A survey evaluating the effects of BOAI after 15 years, indicated the transition from establishing open access as a concept (in 2002) to making open the default [30]. However, this survey showed two important remaining challenges: a) researchers lack meaningful incentives and rewards to openly share their work, and b) lack of sufficient funding to pay for open access-related costs.

There are recent attempts to further formalize the process of sharing in research planning and reporting, such as Transparency and Openness Promotion (TOP) guidelines. These are comprehensive sets of standards for journals, funding bodies, institutions, professional societies, reviewers, and authors [31]. Until now, more than 5000 journals endorsed TOP guidelines.

In addition to researchers sharing their outputs, open science practices assume that, once published, *the content of scientific articles should be free and accessible*. Hence, there is an increasing pressure on publishers to provide open access to publications. For instance, the National Institute of Health (NIH), National Science Foundation (NSF), German Psychological Association demand that data collected as part of publicly funded research are made open to all interested parties. European Commission, as part of innovation funding Horizon 2020 program, made open access to publications mandatory and launched a pilot project to open up publicly funded research data available from 2013 onwards.

C. Incentives

Symbolic rewards. For example, the Center for Open Science proposed that journals appoint *badges* to articles endorsing open practices, i.e., open data, open materials, and pre-registration. Empirical evaluation of this practice demonstrated that introduction of open practices badges had a positive impact of scholars and lead to an increase of data sharing [32].

Employment policies. There is a growing trend of universities, institutes and funding bodies requiring researchers to demonstrate their devotion to open science.

IV. Conclusion

What we presented is not an exhaustive list of proposed measures, rather the measures that were shown to have an effect in raising scientific standards. Open Science movement had numerous positive outcomes. Funding bodies started adopting transparency requirements and supporting financially replication studies. Institutions started providing the scholars with the infrastructure for data sharing, and what is more important, started changing hiring standards and putting more weight to the open science values. Journals are adopting badges to acknowledge open practices, Registered Reports and TOP guidelines.

However, there are still some unresolved issues that we would like to tackle. First is to establish a system of incentives that would encourage good scientific practices and be endorsed by the scholars as sufficiently rewarding. We do not argue that novelty should not be rewarded, on the contrary. We argue that it should not be the only thing that is rewarded. Novelty versus reproducibility is a false dichotomy and goods science requires both. Second challenge draws from the fact that resource allocation is not only required of scientists, it is required of publishers, too. Even though most of them declaratively support the new norms, the current business model is far more lucrative than the open-access model and this still needs to be addressed.

In spite of the challenges, it seems that the current climate is pushing all stakeholders towards more reproducible, credible and accessible science. We can only hope the winds will not turn.

References

[1] K. Popper, *Conjectures and refutations: The growth of scientific knowledge.* London: Routledge and Kegan Paul, 1963

[2] J. P. A. Ioannidis, "Why most published research findings are false," *PLoS Medicine,* vol. 2, pp. E124, 2005

[3] M. R., Munafò, B. A., Nosek, D. V., Bishop, K. S., Button, C. D., Chambers, N. P., du Sert, ... and J. P., Ioannidis, "A manifesto for reproducible science," *Nature Human Behaviour*, vol. 1, pp. 0021, 2017

[4] Open Science Collaboration, "Estimating the reproducibility of psychological science," *Science*, vol. 349, pp. Aac4716, 2015

[5] R. A. Klein, et al., "Many Labs 2: investigating variation in replicability across sample and setting," *Advances in Methods and Practices Psychological Science,* In press

[6] C. F. Camerer, A. Dreber, F. Holzmeister, T. Ho, J. Huber, M. Johannesson, M. Kirchler, G., Nave, B. A. Nosek, T. Pfeiffer, A. Altmejd, N. Buttrick, T. Chan, Y. Chen, E. Forsell, A. Gampa, E. Heikensten, L. Hummer, T. Imai, S. Isaksson, D. Manfredi, J. Rose, E. Wagenmakers, and H. Wu, "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015". *Nature Human Behavior (Letters),* doi: 10.1038.s41562-018-0399-z, 2018

[7] C. F. Camerer, A. Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson,... and E. Heikensten, "Evaluating replicability of laboratory experiments in economics," *Science*, vol. 351, pp. 1433-1436, 2016

[8] C. G. Begley, and J. P. Ioannidis, "Reproducibility in science: improving the standard for basic and preclinical research," *Circulation research*, vol.116, pp. 116-126, 2015

[9] K. Fiedler, and N.Schwarz, "Questionable research practices revisited," *Social Psychological and Personality Science*, vol.7, pp. 45-52, 2016

[10] L. K. John, G. Loewenstein, and D. Prelec, "Measuring the prevalence of questionable research practices with incentives for truth telling," *Psychological Science,* vol.23, pp. 524–532, 2012

[11] B. A. Nosek, J. R. Spies, and M. Motyl, "Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability,"*Perspectives on Psychological Science*, vol. 7, pp. 615-631, 2012

[12] R. Joober, N. Schmitz, L. Annable, and P. Boksa, "Publication bias: what are the challenges and can they be overcome?" *Journal of Psychiatry and Neuroscience,* vol. 37, pp. 149-152, 2012

[13] J. P. A. Ioannidis, "Why science is not necessarily self-correcting," *Perspectives on Psychological Science,* vol. 7, pp. 645–654, 2012

[14] R. MacCoun, and S. Perlmutter, "Blind analysis: hide results to seek the truth," *Nature,* vol. 526, pp. 187–189, 2015

[15] J. E. Gonzales, and C. A. Cunningham, "The promise of pre-registration in psychological research," *Psychological Science Agenda,* 2015. [Online]. Available http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx [Accessed Aug. 20, 2018]

[16] J. Lenzer, J. R. Hoffman, C. D. Furberg, J. P. Ioannidis, and G. P. R. W. Grp, "Ensuring the integrity of clinical practice guidelines: a tool for protecting patients," *British Medical Journal,* vol. 347, pp. F5535, 2013

[17] F. Schwendicke, F., and N. Opdam, "Clinical studies in restorative dentistry: design, conduct, analysis," *Dental Materials,* vol. 24, pp. 29–39, 2017

[18] J. J. Locascio, "Results blind science publishing," *Basic and Applied Social Psychology,* vol. 39, pp. 239-246, 2017

[19] I. L. Žeželj, and B. R. Jokić, "Replication of experiments evaluating impact of psychological distance on moral judgment," *Social Psychology,* vol. 45, pp. 223-231, 2014a

[20] I. L. Žeželj, and B. R. Jokić, "A rejoinder to comments by Eyal, Liberman, & Trope and Gong & Medin," *Social Psychology,* vol. 45, no. 4, pp. 332-334, 2014b

[21] K. S. Button, J. P. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò, "Power failure: why small sample size undermines the reliability of neuroscience," *Nature Reviews Neuroscience*, vol. 14, no. 5, pp. 365, 2013

[22] C. R. Ebersole, et al. "Many Labs 3: evaluating participant pool quality across the academic semester via replication," *Journal of Experimental Social Psychology, 67*, pp. 68–82, 2016

[23] R. A. Klein, K. A. Ratliff, M. Vianello, R. B. Adams Jr, Š. Bahník, M. J. Bernstein,... and Z. Cemalcilar, "Investigating variation in replicability," *Social psychology,* vol. 45, pp. 142–152, 2014

[24] J. F. Landy, M. Jia, I. L. Ding, D. Viganola, W. Tierney, ..., and E. L. Uhlmann, "Crowdsourcing hypothesis tests: Making transparent how design choices shape research results." *Manuscript submitted for publication,* 2018

[25] H. Moshontz, L. Campbell, C. R. Ebersole, H,. IJzerman, H. L. Urry, P. S. Forscher,..., and C. R. Chartier, "The Psychological Science Accelerator: Advancing Psychology through a Distributed Collaborative Network," *Advances in Methods and Practices in Psychological Science*, In press

[26] B.T. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, ... & M. Araya, "LIGO: the laser interferometer gravitational-wave observatory," *Reports on Progress in Physics*, vol 72, pp. 076901, 2009

[27] T. S.Heng, M. W. Painter, K.Elpek, V. Lukacs-Kornek, N. Mauermann, S.J. Turley, ... & S. Davis, "The Immunological Genome Project: networks of gene expression in immune cells," *Nature immunology*, vol 9, pp. 1091-1094, 2008

[28] https://home.cern/ [Accessed Aug. 20, 2018]

[29] G. Gigerenzer, "Statistical Rituals: The Replication Delusion and How We Got There", *Advances in Methods and Practices in Psychological Science*, vol 1, pp. 198-218, 201

[30] N. Shockey, H. Joseph, and M. Hagemann, "BOAI 15 Survey Report," April 12, 2018

[31] B.A. Nosek, G. Alter, G.C. Banks, D. Borsboom, S. Bowman, S.J. Breckler, ... & M. Contestabile, "Promoting an open research culture", *Science*, vol 348, pp. 1422-1425, 2015.

[32] M. C. Kidwell, L. B. Lazarevic, E. Baranski, T. E., Hardwicke, S. Piechowski, L-S. Falkenberg, C. Kennett, A. Slowik, C. Sonnleitner, C. Hess-Holden, T. M. Errington, S. Fiedler, and B. A. Nosek, "Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency," *PLoS Biology,* vol. 14, pp. 1-15, 2016