Quality & Quantity (2022) 56:1597–1609
https://doi.org/10.1007/s11135-021-01196-6

Improving the reproducibility of findings by updating research methodology

Joseph Klein[1]

Accepted: 26 June 2021 / Published online: 8 July 2021

Abstract
The literature discusses causes of low reproducibility of scientific publications. Our article adds another main cause—uncritical adherence to accepted research procedures. This is evident in: (1) anachronistically requiring researchers to base themselves on theoretical background even if the studies cited were not tested for reproducibility; (2) conducting studies suffering from a novelty effect bias; (3) forcing researchers who use data mining methods and field-based theory, with no preliminary theoretical rationale, to present a theoretical background that allegedly guided their work—as a precondition for publication of their findings. It is possible to increase research validity in relation to the above problems by the following means: (1) Conducting a longitudinal study on the same participants and only on them; (2) Trying to shorten the time period between laboratory experiments and those on humans, based on cost–benefit considerations, anchored in ethical norms; (3) Reporting the theoretical background in a causal modular format; (4) Giving incentives to those who meet the above criteria while moderating the pressure for fast output.

Keywords Research methods · Novelty effect · Replication · Meta-analysis

1 Literature review

Replication of research is considered one of the basic criteria for determining its quality (Pulverer 2015; Van Bavel et al. 2016). The low percentage of studies in the various disciplines that meet the replication test in its various meanings as described below (Baker 2016; Ioannidis 2005) helps to explain the difficulty faced by the authors of articles in presenting synergistic models that incorporate cumulative research findings over time. The literature addresses several specific factors that explain the low replication rate and offers solutions. Our article attributes part of the problem to research procedures that have not been updated based on current information, and offers options for reducing the extent of the reproducibility problem through such updating.

Scientific research is like an inverted pyramid. One generation passes the knowledge it has acquired to the next, to assist future generations in their efforts to add new layers

✉ Joseph Klein
Joseph.Klein@biu.ac.il

[1] School of Education, Bar-Ilan University, 52900 Ramat-Gan, Israel

of insight to present knowledge (Flexner 2017). The underlying assumption of this intellectual project is that the preceding layers of knowledge are stable because they have met the criterion of scientific review (Pulverer 2015). One of the most basic criteria for research validity is reproducibility (Caplan and Redman 2018; Simons 2014; Van Bavel et al. 2016). Several types of replication tests exist: Close replication, Constructive replication, Conceptual replication in the laboratory, Conceptual replication in the field and repeatability (Hüffmeier et al. 2016). One meaning of these terms is the ability to recreate the results of observations or experiments published in the scientific literature (reproducible research). The second meaning is the ability to conduct an experiment or observation again and to obtain the same results time after time (replicability). In medicine and psychology, the term repeatability usually refers to the extent of similarity in results of repeated tests conducted on the same participant and in certain cases on different participants. The common denominator linking the three is that researchers do not base their conclusions on one-time measurement and analysis of their findings. In this article we focus mainly on the two last meanings, but also make reference to the first.

Studies published without this test over long periods of time lack a basic touchstone for assessing their credibility and are thus insignificant to science (Bishop. 2020; Munafò et al. 2020; Popper 2005). In 1995, the negation of findings that did not meet the test of reproducibility spurred physicist Prof. Tzvi Lipkin of the Weizman Institute of Scientific Research and Prof. Alexander Cohen, Director of the Israel Institute for Biological Research (IIBR) to publish a humorous journal for scientists called the Journal of Irreproducible Results, or JIR for short.

From the literature it is apparent that many scientific journals and conferences approve the publication or presentation of short-term studies of limited reproducibility rating (Nosek et al. 2015; Resnik and Shamoo 2017; Munafò et al. 2020; Begley and Ellis 2012), without noting this methodological limitation. This is evident in disciplines such as economics (Anderson et al. 2008), political science (King 1995), genetics, pharmacology, biology (Begley and Ellis 2012; Kaplan and Irvin 2015; Ioannidis 2005), and oncology (Pusztai et al. 2013).

One of the triggers, although not the first, for questioning the validity of research in the social sciences was the finding that for about 64 of every 100 studies published in select psychology journals in 2008, later repetition of the experiments or measures yielded results that differed from those published in 2008 (Nosek et al. 2015). In the last decade a broad-scale Reproducibility Project has been carried out in conjunction with universities and research institutes, intended to re-examine the validity of several central theories in psychology. The project attempts to replicate the findings of the original studies that served as a basis for the theories in question (Nosek et al. 2012).

2 Causes of the reproducibility problem

The problem of reproducibility, in terms of repeated measurements at a later date, is explained as difficulty in planning conditions that are completely identical to the original ones for replicating measurements of a study (Ferguson and Heene 2012; John et al. 2012), and contextual differences in various measurements such as cultural parameters and interpersonal interactions (Gilbert et al. 2016; Rousseau and Fried 2001; Schweinsberg et al. 2016). Some attribute the reproducibility problem to research reliance on small samples characterized by low statistical power, such as the small size effect (Schmidt and Oh 2016).

Insufficient control of the central factors that explain the connection between variables raises the possibility that inconsistent findings will be obtained in repeat measurements (Bakker et al. 2012). The problem of lack of reproducibility is also attributed to insufficiently strict adherence by journal editorial boards to the methodological level of the studies, fueled by a fervor to publish surprising results (Schmidt and Oh 2016; Yong 2012; Young et al. 2008). At the same time, those writing replication studies have little chance of their articles being accepted for publication. The reason often given by editorial boards is that findings similar to previous ones do not provide sufficient innovation (Makel and Plucker 2014). The chances are also low of publishing an article about failure to confirm a hypothesis that arises from the literature (Ioannidis 2005). Examination of the literature reveals other important factors that may impair reproducibility, but they are not mentioned in discussions of the problem.

3 Additional causes of the reproducibility problem and the difficulty of publishing findings with reproducibility data

3.1 The novelty effect

Bishop (2020) surveys a broad range of cognitive biases that impinge on the validity of researchers' work. We will examine one of these biases in depth, the novelty effect (Thorgusen et al. 2016). The novelty effect is defined as a cognitive, affective and/or behavioral change that occurs as a result of initial exposure of study participants to a new situation, such as an experiment or a new learning technology. This change eventually wanes or disappears as participants acclimate to the new situation. Thus, the functioning of individuals during their initial exposure to an experiment should not be used to determine their expected functioning at a later stage or over time (Gravetter and Forzano 2018). The bias is evident among humans (Belton and Sugden 2018) and animals (Domjan 2018). It is also connected to the Hawthorne effect (Rosema et al. 2011) and to the placebo effect, in which psycho-neurological mechanisms affect the experiences of individuals, at times only at the symptomatic level (Ashar et al. 2017; Benedetti 2013).

It is easy to control for this bias when the experimental effect and the novelty effect are not interdependent and each one has an independent effect on the result. Measurement of the experimental effect is performed by comparing the results of the experimental group with those of the control group that received a placebo. Many studies are based on the assumption of non-dependence between the two factors. However, the extensive research that connects the efficacy of medical treatment to the psychological and mental condition of patients raises an alternative possibility, namely that of an interaction between experimental effect and novelty effect in the experimental group. Such interaction can increase or reduce the intensity of both effects on the result, beyond the individual effect of each factor by itself. In an interactive condition, calculating the size of the experimental effect based on the assumption of non-dependence is irrelevant. Measuring the net effect of the experimental effect requires long term monitoring, until the bias (novelty effect) dissipates. A long-range replication study conducted on different participants each time exposes each new group of participants to the novelty effect and thus loses validity. The novelty effect is expected to be expressed in laboratory experiments and in various interventions, both physical and psychological. In contrast, in studies based on non-participatory observations and on questionnaires, it is doubtful that the novelty effect is relevant.

3.2 Limiting the number of researchers involved in published articles

3.2.1 Withholding academic accreditation for multi-partner research

Comprehensive, long-term studies with repeated measures often require the involvement of many researchers (Frith 2020; Poldrack 2019). A study headed by Nosek included 270 researchers, and without such broad cooperation, the findings obtained would have been much more limited (Nosek et al. 2015). The need to involve experts from various disciplines in some studies (Munafò et al. 2017; Santos et al. 2017; Gil et al. 2015) also increases the number of research partners. The probability of publishing a study with many participating researchers is limited by several factors. Some university faculties limit the number of joint articles that are recognized for accreditation towards academic promotion. For example, the guidelines for academic promotion in the social sciences at one university states that researchers seeking promotion are requested to indicate the number of articles in which they appear as the lead author and the number articles in which they appear as second author. Third place or lower does not appear in the document at all. Moreover, the measure of academic success requires publication of a minimum number of studies every year (Pulverer 2015). The rapid pace of publication and the research burden this entails (Frith 2020) do not leave sufficient time to initiate complex, multi-researcher studies. The limitations on delving deeply enough into a subject imposed by these circumstances increase the likelihood of compromising reproducibility (Ioannidis 2005). From a regulatory viewpoint, it is simpler to supervise the quality of researchers' work by means of uniform easily accessible measures. These criteria are roundly criticized in the literature but the fact that they have still not been updated indicates that research forces must be pooled in order to assess and promote desired changes. The section on proposed solutions to increase the reproducibility of research and alternatives to this test will discuss possible directions for a solution.

3.2.2 The difficulty of publishing studies that meet the reproducibility test, because of the use of innovative methodologies

The prevailing structure of articles appearing in scientific journals today includes a literature review, hypotheses and/or research questions arising from the review, and details of the methodology used for examining them. This report format reflects how some findings are obtained. However, other findings are revealed by means of different processes such as data mining and grounded theory. The starting point for researchers who utilize these procedures is not necessarily preliminary knowledge of the literature but rather systematic sleuthing to discover knowledge embedded in information bases, or to identify consistent knowledge discerned by personnel in the field whose reports did not appear in the literature. While some data mining research papers are in fact based on a preliminary literature review, this does not generalize to all of them. One of the long-term retrospective advantages of the data mining method is its potential for identifying knowledge that meets the test of reproducibility. This is because it examines big data, i.e., large and multi-annual data instead of samples whose representative nature is at times questionable (Linden and Yarnold 2016a, b). The grounded theory method draws on insights from field experiences, usually over a long period, rather than one-time phenomena (Glaser and Strauss 2017), which means that they too are potentially reproducible. The data mining method

is gradually becoming more common and highly central in a number of disciplines. In medical research it is considered to have great potential for identifying causes and means of treating or preventing diseases, especially by pooling and cross-referencing data about known mutations, using data bases of clinical studies, medical reports, reports on reactions to treatments, symptoms and the like (Johannes 2016). At the same time, caution is needed so as not to overgeneralize conclusions from big data. The population at large encompasses many subgroups and individuals with their own unique characteristics. In light of this insight, the medical profession is seeking to develop personally tailored medicine, adapted to the unique traits and needs of each individual (Ventegodt 2016).

Researchers who obtain reproducible findings by means of these methods, without reference to any preliminary theoretical basis, will have difficulty publishing their papers in the authentic format based on the actual process that they followed. If they adapt their papers to suit the classical format, with a literature background that allegedly served as a platform for their work, they are actually subverting the truth by not reporting their authentic work process. Moreover, such false reports will make it more difficult for other researchers to replicate the work. Based on this alleged theoretical background, they may deduce that other or additional variables should be examined.

It should be emphasized that the use of new methodologies does not guarantee research validity. It is possible for significant mistakes to be made while analyzing and drawing conclusions based on data mining. This is especially true when broad data bases containing many different populations and variables are used, making the process highly complex. Researcher bias may affect the choice of variables to be investigated and those to be disregarded (Johannes 2016; Leek and Peng 2015). Therefore, researchers emphasize the need to train a generation of researchers will high awareness of what constitutes high level research using advanced data processing skills.

3.2.3 Proposed solutions to increase the reproducibility of research

One of the solutions suggested for improving reproducibility is to raise researchers' awareness of the importance of planning their work with strict adherence to procedures (Funder et al. 2014; Pulverer 2015). It has also been suggested that the method of rewarding researchers be altered: instead of recompense based on results it should be based on diligence in carrying out the research, regardless of results or of frequency of publication (Nosek et al. 2012). Similarly, it has been recommended to require full transparency for the entire research process, so that all researchers can examine the quality of the research methodology and thus to lay the basis for their own replication of the study (Brandt, 2014; Dreber et al. 2015). In clinical cardiac studies (Ioannidis 2005), the demand for full transparency in the research procedure contributed to a reduction in the report of positive results. It was therefore proposed to limit the rate of publication required from researchers, to allow them to conduct more thorough work (Ioannidis 2005), including replication of studies (Maxwell et al. 2015; Saey 2015). To assist researchers who wish to conduct replication studies, archives of research data were established so that they data could be tested. In this vein, the editorial boards of many journals ask potential contributors of articles if they are willing to share their research data with the community at large.

The difficulties entailed in carrying out replication studies have spurred a search for alternatives, such as meta-analytical studies. Such studies do not contain replicated measurements but they are based on conclusions drawn from a large number of papers each of which may be limited in scope by itself, but taken together constitute a broad sample.

Analysis of comparative findings from many studies from around the world provides knowledge about the extent of repetition of results for the same measurements (Schmidt and Oh 2016). Using meta-analysis, it is possible to control for the size of the effect being measured, statistical power, sample size, degrees of freedom and missing values. Successful replication of the initial meta-analyses by repeated examinations of the same original data bases by researchers with different perspectives also increases validity (Lakens et al. 2016). Despite the advantages of this alternative, meta-analyses based on data bases composed of short-term rather than longitudinal studies are subject to the difficulty of isolating the size of the experimental effect from the novelty effect, where an interaction exists between them.

Other easier alternatives have been proposed for measuring reproducibility, among them using triangulation to amass data (Munafò and Smith 2018), crowdsourced testing of a research hypothesis, including the cross-referencing of data from independent studies (Landy et al. 2020), exploratory and confirmatory factor analysis studies, obligatory detailed recording of the research process and full transparency for all those wishing to examine it (Allen and Mehler 2019; Field et al. 2020). The stringency demonstrated in these approaches increases measurement validity but at the same time may entail exposure to the novelty effect, if the data bases are composed solely of short-term measurements.

We believe that beyond the basic recommendations in the literature, even greater potential exists for improving reproducibility, and the means to this end are described in the following section.

3.3 Additional solutions for increasing reproducibility

3.3.1 Re-assessment of the ethical principles in research

Conducting an experiment in the authentic environment in which the phenomenon under study occurs creates optimal conditions for testing reproducibility. Prevailing laws and ethics permit authentic testing only after strictly controlled laboratory tests have been conducted. Such testing is usually prolonged and intended to prevent harm to the participants, especially, but not only, if the target population is composed of humans. Because of the advantage of experimenting in an authentic environment it is necessary to constantly examine possible options for shortening the time needed for laboratory experimentation. To this end, two questions should be discussed, one empirical and the other ethical.

3.3.2 The empirical question

How would the long- and short-term cost–benefit relationship be affected if controlled sample experiments were conducted in an authentic environment before all laboratory tests were completed? The early authentic environment option should be compared to the results of authorizing authentic experimentation only after conclusion of the entire laboratory process. The consequences of the two options can be illustrated by the case of life-saving pharmacological drugs (medications). In the short term, administering experimental drugs to a limited sample of subjects before laboratory tests have been fully completed will almost certainly cause a significantly greater number of negative than beneficial effects on participants, since only a small percentage of the substances tested in the laboratories is eventually approved for use in humans. Most substances are rejected either because they are ineffective or because they have a negative effect (Pavlou et al. 2013). Such harm as

they cause may engender short-term problems, prolonged impairment of quality of life or even shortened life span. On the other hand, accelerating the testing procedure in authentic conditions may lead to a more rapid discovery of medications and drugs with positive potential, with the long-term result of effective medication reaching a broader target population sooner than if the drugs were released only after conclusion of the laboratory experiments. Empiric investigation to assess help-harm proportions when using the two constructs for developing drugs requires specific test case studies. Even in this case, it is possible that findings will vary substantially from case to case. If in the long-term view the cost–benefit relationship indicates that human lives will be saved by conducting authentic experimentation and reproducibility tests earlier in the process, this will pave the way to addressing the normative aspect of the issue.

The basic ideas of this article were formulated before the Covid 19 era. In the interim, global experience with the pandemic has shown the vital importance of combining strict adherence to meticulous research processes with a significant condensing of the development process by means of experimentation on volunteers, which may save many lives in the future. To date, it is not known whether the accelerated development process and the rapid transition to testing on human beings will be a one-time event limited to this vaccine and accepted norms will continue to dictate stringent long-term experimentation in the future. Alternatively, the experience in developing a vaccine against Covid 19 may inspire a willingness to include experimentation on human volunteers earlier in the process, based on assessments that such action will prevent the loss of many lives.

In discussing this issue two different perceptions should be kept in mind. One is the general attitude toward people, animals or plants participating in experiments and the harm caused to them during such participation. Their suffering is evident and their death as a result of the experiment is mourned vociferously. The second perception pertains to lives that can be saved in the future, which are viewed as statistics that do not arouse strong feelings or compassion.

Changing the process such that authentic experiments, including reproducibility tests, come earlier in the process may save human lives, and pave the way for clarifying the normative aspect of the issue.

3.3.3 The normative question

The assumption that conducting an experiment in the authentic environment in which a phenomenon occurs creates optimal conditions for testing reproducibility, raises an important question: Is it legitimate, normatively speaking, to allow a few volunteers to endanger themselves by testing a medication whose effectiveness has not yet been proven in the laboratory, in order to save many other lives? Dark regimes use prisoners to this end. Significant endangerment of life for the sake of other people is a common characteristic of the military, whose volunteers (or conscripts) take part in battles, some of which levy a heavy price in casualties, in order to save others. These soldiers do not save the lives only of citizens of their home country. The participation of the United States Army in Europe in the two World Wars represented a norm in which the lives of soldiers were willingly endangered for the sake of other nations. These actions surely served US national interests although most residents of the United States faced no existential threat and the country's territorial boundaries were not in immediate jeopardy. It is worth examining whether volunteering to the point of endangerment of life in order to prevent violence by a human enemy is different from volunteering to the point of endangerment of life in order to limit

serious damage caused by sickness factors originating in nature. If the answer to this normative question is that volunteering is equally worthy in both domains, it will be necessary to clarify its operational values. One of the expected side effects of approving early authentic experimentation in certain circumstances is the acceleration of research to monitor and neutralize the toxicity of substances during experiments.

3.4 Updating the report format of articles in journals

3.4.1 Literature review in modular format that emphasizes studies meeting the reproducibility test

Readers of articles will benefit from reading a literature review that contains a concise and descriptive graphic model of the main causal variables involved—direct and indirect, emphasizing the paths that were proven to meet—or not meet—the test of reproducibility. It is possible to append to this a short description of the researchers' rationale for selecting the variables presented.

The advantage of presenting existing knowledge in concise modular form has been discussed in the literature (Castellani 2014) and is also reflected in researchers' comments in specific works they have carried out. In a study on motives for alcohol consumption and its dangers, the authors note that comprehensive understanding of the phenomenon requires prior paradigmatic and systemic familiarity of the issue (Apostolopoulos et al. 2018). In a study of Hispanic immigrants, criticism was leveled against researchers who relied on limited and non-systemic knowledge, as a result of which their contribution to an understanding of the risk factors facing immigrants employed in hard labor at low wages was limited (Sönmez et al. 2017). There are many ways to present causal formational knowledge. Findings have accumulated about the advantages of its visual presentation on a continuum, in a figure or a graph (Lu et al. 2020; Wang and Mueller 2015).

Joint publication of causal models on the internet that meet the reproducibility test, encompassing the findings of many studies, may lead to maximal utilization of causal information and serve as a basis for research in the literature. Allowing researchers to use an identically formulated literature background is not something to be taken for granted. Some researchers and journal editorial boards require all articles to contain a uniquely formulated literature review that does not cite a list of literature insights that appeared in other studies. The author of the present article sent an article for review that was returned to the author by the journal board with a letter stating that he had plagiarized material. Appended to the rejection letter was a report produced by software devised to identify articles copied from the internet. It found an identical 30% overlap between the literature review in the submitted article and that in another article. Examination of the report revealed that the "plagiarized" part was self-plagiarism, as the author had copied from another article that he himself had written. The report did not find any plagiarism of the research question, methodology, findings or discussion. The editorial board's attention was brought to these data but to no avail. The editors contended that the author was required to provide unique material in the literature review for each of his articles. The main contribution or uniqueness of researchers is not reflected in the review of knowledge reported by others who preceded them but rather in the new information that they intend to add. Moreover, having to write a new and unique literature review for each article forces researchers to invest thought and time in a part of the study that does not contribute to the essence of its innovation.

Furthermore, some researchers who excel in conducting studies may not be endowed with the ability to write a unique literature review for each paper they produce.

4 Discussion

Reproducibility in its many meanings, is not the be-all and end-all criterion. Conducting an initial experiment without passing this test may provide knowledge about the feasibility of researching a subject more deeply. The level of reproducibility of research findings should be judged on a continuum. High correspondence between measurements over time should not be taken as a categorical definition of success while any other condition is considered failure. Partial but consistent correspondence between replicated measurements of experimental drug/pharmacological treatments may be perceived as a significant improvement in quality of life by study participants. In psychological experiments, such findings may indicate an improvement in learning effectiveness and a slight albeit significant improvement in achievements. What is needed are research tools that strike a balance between very strict research requirements such as reproducibility, and the need to promote research within the constraints of existing financial and temporal resources.

The literature discusses the factors that cause problems of reproducibility and concrete means of improving the situation (Funder et al. 2014; Pulverer 2015). Our article indicates an additional cause of the reproducibility problem: insufficient distinction in research between findings exposed to the novelty effect bias and those studies free of this shortcoming. Even reproducible long-term studies may suffer from this bias in conditions in which they are conducted on new participants each time. Paradoxically, the requirement to test the validity of findings on new participants each time, as in exploratory and confirmatory analysis, creates a significant ancillary problem by exposing new participants to the novelty effect each time. A possible solution to this problem is to conduct longitudinal studies with a number of comparison groups, among them those that receive treatment, others a placebo and others nothing. The fact that monitoring continues over time will help to moderate or dissipate the novelty effect.

Because of their complexity, reproducible studies require cooperation among researchers, (Bishop 2020; Frith 2020; Poldrack 2019). The average number of cooperating authors of articles has risen over the years (Haws et al. 2018). An accurate comparative examination of the various disciplines is complex, because those bibliometric investigations that have been published were conducted years apart. The average number of authors per article varies from one field of study to another, but all are more than one. The average number of cooperating authors in medical articles in the Journal of Arthroplasty rose from 3.54 in 1986 to 4.98 in 2015 (Lehman et al. 2017). Librarianship and the information sciences reported 2.27 authors per article (Verma and Singh 2017), and psychology 1.67 (Zafrunnisha and Pulla Reddy 2009). The number of authors is influenced not only by the desire for complex examination and multidimensionality of each issue under study, but also by the need to publish or perish (Génova and de la Vara 2019; Kiai 2019), which stimulates the creation of partnerships so as to increase the number of publications.

The integration of experts from diverse fields expands the interdisciplinary perspective of the examination and enhances the probability of the findings meeting the test of reproducibility.

Multi-author planning encounters difficulties because the criteria for academic promotion in some academic institutions and in some disciplines, such as the social sciences,

incentivize the lead author. Less weight is given to the second author, and the third author and downward are barely considered, based on the assumption that their contribution is negligible. Empirical findings indicate that in most multi-author articles, all the authors usually contribute noteworthy information, although the contribution of the first and last authors in the list is considered greater than the others (Corrêa et al. 2017; Larivière et al. 2016). At the same time, a multiplicity of authors raises the potential for professional and ethical dilemmas, as in the following scenario. A scientist initiates and plans a study in the area of his expertise. Implementation is assigned to experts hired for the various stages of the work. These include experts in information management to write the theoretical background, methodological experts to plan the research design, select the measurement tools and characterize the study sample. A group of experimenters is appointed to conduct the experiments, expert statisticians analyze the findings and an experienced scientific editor is hired to write the research report in article format. An integrator coordinates the activity among all the experts. The researcher who initiated the study receives detailed updates of progress throughout the process and thus ensures that it is being carried out according to his plan, even though he is not actively involved in its implementation. This researcher reviews the final article and approves it for publication. The question now is who of those involved as initiator and implementers should be included in the list of article authors and in which order.

Despite the advantages of the current research quality measures, the limitations ascribed to them underscore the need to seek ways to update them, and the literature offers possible directions. Perhaps academic accreditation could also be awarded to research that is recognized as being of great value and as adhering to the most stringent methodology, even if its results do not provide statistically significant support to the research hypotheses. The contribution of such research is twofold: a. it negates directions of thought that were considered significant, and b. it encourages researchers to conduct their work intrepidly and without trying to bias their findings lest disproval of their hypotheses be held against them. Such an approach requires finding worthy platforms willing to publish such research, as the editors of many journals might not accept such articles.

Despite criticism of the large number of studies that do not pass the reproducibility test, whether because replication results differ completely from the original or because reproducibility was not tested at all, many studies do meet this criterion annually. They are published in a large number of journals ranked high by academic standards and produce extensive knowledge. As a result, few question the great progress science has made, as reflected in increasing life span and improved quality of life in many domains. The gap between real and ideal reproducible findings can be perceived as a catalyst for continuing the search for ways to reduce this gap. The world is rich in researchers with excellent abilities and their output will improve even more when, from time to time, they put their heads together to test the validity of how they work, in order to update it. This is a subject about which wise men in the past have said: Sometimes a shortcut is the longer way and the long way is the shortest.

References

Anderson, R.G., Greene, W.H., McCullough, B.D., Vinod, H.D.: The role of data/code archives in the future of economic research. J. Econ. Methodol. **15**(1), 99–119 (2008)

Allen, C., Mehler, D.M.: Open science challenges, benefits and tips in early career and beyond. PLoS Biol. **17**(5), e3000246 (2019)

Apostolopoulos, Y., Lemke, M.K., Barry, A.E., Lich, K.H.: Moving alcohol prevention research forward—Part I: introducing a complex systems paradigm. Addiction **113**(2), 353–362 (2018)
Ashar, Y.K., Chang, L.J., Wager, T.D.: Brain mechanisms of the placebo effect: an affective appraisal account. Annu. Rev. Clin. Psychol. **13**, 73–98 (2017)
Baker, M.: 1,500 scientists lift the lid on reproducibility. Nat. News **533**(7604), 452–454 (2016)
Bakker, M., van Dijk, A., Wicherts, J.M.: The rules of the game called psychological science. Perspect. Psychol. Sci. **7**, 543–554 (2012)
Begley, C.G., Ellis, L.M.: Drug development: raise standards for preclinical cancer research. Nature **483**(7391), 531–533 (2012)
Benedetti, F.: Placebo and the new physiology of the doctor-patient relationship. Physiol. Rev. **93**(3), 1207–1246 (2013)
Belton, C.A., Sugden, R.: Attention and novelty: an experimental investigation of order effects in multiple valuation tasks. J. Econ. Psychol. **67**, 103–115 (2018)
Bishop, D.V.: The psychology of experimental psychologists: overcoming cognitive constraints to improve research: the 47th Sir Frederic Bartlett Lecture. Q. J. Exp. Psychol. **73**(1), 1–19 (2020). https://doi.org/10.1177/1747021819886519
Brandt, M.J., IJzerman, H., Dijksterhuis, A., Farach, F.J., Geller, J., Giner-Sorolla, R., Van't Veer, A.: The replication recipe: What makes for a convincing replication? J. Exp. Soc. Psychol. **50**, 217–224 (2014)
Caplan, A.L., Redman, B.K. (eds.): Getting to Good: Research Integrity in the Biomedical Sciences. Springer (2018)
Castellani, B.: Focus: complexity and the failure of quantitative social science. Focus **12**, 12 (2014)
Corrêa, E.A., Jr., Silva, F.N., Costa, L.D.F., Amancio, D.R.: Patterns of authors contribution in scientific manuscripts. J. Informet. **11**(2), 498–510 (2017)
Domjan, M.: Introduction to food neophobia: historical and conceptual foundations. In: Reilly, S. (ed.) Food Neophobia: Behavioral and Biological Influences, pp. 15–30. Elsevier, Amsterdam (2018)
Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B.A., Magnus Johannesson, M.: Using prediction markets to estimate the reproducibility of scientific research. Proc. Natl. Acad. Sci. **112**(50), 15343–15347 (2015)
Ferguson, C.J., Heene, M.: A vast graveyard of undead theories publication bias and psychological science's aversion to the null. Perspect Psychol Sci **7**(6), 555–561 (2012)
Field, S.M., Wagenmakers, E.J., Kiers, H.A., Hoekstra, R., Ernst, A.F., van Ravenzwaaij, D.: The effect of preregistration on trust in empirical research findings: results of a registered report. Royal Society Open Science **7**(4), 181351 (2020)
Flexner, A.: The Usefulness of Useless Knowledge. Princeton University Press (2017)
Funder, D.C., Levine, J.M., Mackie, D.M., Morf, C.C., Sansone, C., Vazire, S., West, S.G.: Improving the dependability of research in personality and social psychology: recommendations for research and educational practice. Personal. Soc. Psychol. Rev. **18**, 3–12 (2014)
Frith, U.: Fast lane to slow science. Trends Cogn. Sci. **24**(1), 1–2 (2020)
Haws, B.E., Khechen, B., Movassaghi, K., Yom, K.H., Guntin, J.A., Cardinal, K.L., Singh, K.: Authorship trends in Spine publications from 2000 to 2015. Spine **43**(17), 1225–1230 (2018)
Génova, G., de la Vara, J.L.: The problem is not professional publishing, but the publish-or-perish culture. Sci. Eng. Ethics **25**(2), 617–619 (2019)
Johannes, M.: Big Data for Big Pharma: An Accelerator for the Research and Development Engine?, vol. 19. ibidem-Verlag/ibidem Press (2016a)
Gilbert, D.T., King, G., Pettigrew, S., Wilson, T.D.: Comment on "Estimating the reproducibility of psychological science." Science **351**(6277), 1037–1037 (2016)
Gill, S.V., Vessali, M., Pratt, J.A., Watts, S., Pratt, J.S., Raghavan, P., DeSilva, J.M.: The importance of interdisciplinary research training and community dissemination. Clin. Transl. Sci. **8**(5), 611–614 (2015)
Glaser, B.G., Strauss, A.L.: Discovery of Grounded Theory: Strategies for Qualitative Research. Taylor and Francis, London (2017)
Gravetter, F.J., Forzano, L.A.B.: Research Methods for the Behavioral Sciences. Cengage Learning (2018)
Hüffmeier, J., Mazei, J., Schultze, T.: Reconceptualizing replication as a sequence of different studies: a replication typology. J. Exp. Soc. Psychol. **66**, 81–92 (2016)
Ioannidis, J.P.: Why most published research findings are false. PloS Med **2**(8), e124 (2005)
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)
Kaplan, R.M., Irvin, V.L.: Likelihood of null effects of large NHLBI clinical trials has increased over time. PLoS ONE **10**(8), e0132382 (2015)

Kiai, A.: To protect credibility in science, banish "publish or perish." Nat. Hum. Behav. **3**(10), 1017–1018 (2019)
King, G.: Replication, replication. PS Polit. Sci. Polit. **28**(3), 444–452 (1995)
Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., Sugimoto, C.R.: Contributorship and division of labor in knowledge production. Soc. Stud. Sci. **46**(3), 417–435 (2016)
Leek, J.T., Peng, R.D.: Opinion: reproducible research can still be wrong: adopting a prevention approach. Proc. Natl. Acad. Sci. **112**(6), 1645–1646 (2015)
Lehman, J.D., Schairer, W.W., Gu, A., Blevins, J.L., Sculco, P.K.: Authorship trends in 30 years of the Journal of Arthroplasty. J. Arthroplasty **32**(5), 1684–1687 (2017)
Linden, A., Yarnold, P.R.: Using data mining techniques to characterize participation in observational studies. J. Eval. Clin. Pract. **22**(6), 839–847 (2016a). (**This was in Hebrew but not English**)
Lu, Y., Meisami, A., Tewari, A., Yan, W.: Regret analysis of Bandit problems with causal background knowledge. In: Conference on Uncertainty in Artificial Intelligence, pp. 141–150. PMLR (2020)
Landy, J.F., Jia, M.L., Ding, I.L., Viganola, D., Tierney, W., Dreber, A., Ly, A.: Crowdsourcing hypothesis tests: making transparent how design choices shape research results. Psychol. Bull. **146**(5), 451–479 (2020). https://doi.org/10.1037/bul0000220. (**Epub 2020 Jan 16**)
Lakens, D., Hilgard, J., Staaks, J.: On the reproducibility of meta-analyses: six practical recommendations. BMC Psychol. **4**(1), 24 (2016)
Linden, A., Yarnold, P.R.: Using data mining techniques to characterize participation in observational studies. J. Eval. Clin. Pract. **22**(6), 839–847 (2016b)
Makel, M.C., Plucker, J.A.: Facts are more important than novelty: replication in the education sciences. Educ. Res. **43**(6), 304–316 (2014)
Munafò, M.R., Smith, G.D.: Robust research needs many lines of evidence. Nature **553**, 399–401 (2018)
Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ioannidis, J.P.: A manifesto for reproducible science. Nat. Hum. Behav. **1**(1), 21 (2017)
Munafò, M.R., Chambers, C.D., Collins, A.M., Fortunato, L., Macleod, M.R.: Research culture and reproducibility. Trends Cogn. Sci. **24**(2), 91–93 (2020)
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**(6), 487–498 (2015)
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)
Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T.A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T.: Promoting an open research culture. Science **348**(6242), 1422–1425)2015)
Pavlou, M.P., Diamandis, E.P., Blasutig, I.M.: The long journey of cancer biomarkers from the bench to the clinic. Clin. Chem. **59**(1), 147–157 (2013)
Poldrack, R.A.: The costs of reproducibility. Neuron **101**(1), 11–14 (2019)
Popper, K.: The Logic of Scientific Discovery. Routledge (2005)
Pulverer, B.: Reproducibility Blues. EMBO J. **34**(22), 2721–2724 (2015)
Pusztai, L., Hatzis, C., Andre, F.: Reproducibility of research and preclinical validation: problems and solutions. Nat. Rev. Clin. Oncol. **10**(12), 720 (2013)
Resnik, D.B., Shamoo, A.E.: Reproducibility and research integrity. Account. Res. **24**(2), 116–123 (2017)
Rosema, N.A., Hennequin-Hoenderdos, N.L., Berchier, C.E., Slot, D.E., Lyle, D.M., van der Weijden, G.A.: The effect of different interdental cleaning devices on gingival bleeding. J. Int. Acad. Periodontol. **13**(1), 2–10 (2011)
Rousseau, D.M., Fried, Y.: Location, location, location: contextualizing organizational research. J. Organ. Behav. **22**(1), 1–13 (2001)
Saey, T.H.: Repeat performance. Sci. News **187**, 21–26 (2015)
Santos, P.V.C., Oliveira, S.R.I., Jezus, S.V.: Perception and performance of the interdisciplinary team In serving drug dependents in specialized network of mental health, Sinop, Mato Grosso. Sci. Electron. Arch. **10**(4), 81–86 (2017)
Schmidt, F.L., Oh, I.S.: The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? Arch. Sci. Psychol. **4**(1), 32 (2016)

Schweinsberg, M., Madan, N., Vianello, M., Sommer, S.A., Jordan, J., Tierney, W., Srinivasan, M.: The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. J. Exp. Soc. Psychol. **66**, 55–67 (2016)

Simons, D.J.: The value of direct replication. Perspect. Psychol. Sci. **9**, 76–80 (2014)

Sönmez, S., Apostolopoulos, Y., Lemke, M.K., Hsieh, Y.C.J., Karwowski, W.: Complexity of occupational health in the hospitality industry: dynamic simulation modeling to advance immigrant worker health. Int. J. Hosp. Manag. **67**, 95–105 (2017)

Thorgusen, S.R., Suchy, Y., Chelune, G.J., Baucom, B.R.: Neuropsychological practice effects in the context of cognitive decline: contributions from learning and task novelty. J. Int. Neuropsychol. Soc. **22**(4), 453–466 (2016)

Van Bavel, J.J., Mende-Siedlecki, P., Brady, W.J., Reinero, D.A.: Contextual sensitivity in scientific reproducibility. Proc. Natl. Acad. Sci. **113**(23), 6454–6459 (2016)

Ventegodt, S., Kandel, I., Ervin, D.A., Merrick, J.: Concepts of holistic care. In: Health Care for People with Intellectual and Developmental Disabilities across the Lifespan, pp. 1935–1941. Springer, Cham (2016)

Verma, N., Singh, K.: Authors productivity and degree of collaboration in journal of librarianship and information science (JOLIS) 2010–2016. Int. J. Libr. Inf. Stud. **7**(4), 1–6 (2017)

Wang, J., Mueller, K.: The visual causality analyst: an interactive interface for causal reasoning. IEEE Trans. vis. Comput. Graph. **22**(1), 230–239 (2015)

Yong, E.: Bad copy. Nature **485**(7398), 298 (2012)

Young, N.S., Ioannidis, J.P., Al-Ubaydli, O.: Why current publication practices may distort science. PLoS Med. **5**(10), e201 (2008)

Zafrunnisha, N., Pulla Reddy, V.: Authorship pattern and degree of collaboration in psychology. Ann. Libr. Inf. Stud. **17**(1), 255–261 (2009)