# Responding to the growing issue of research reproducibility

Donald R. Baer, and Ian S. Gilmore

Citation: Journal of Vacuum Science & Technology A **36**, 068502 (2018); doi: 10.1116/1.5049141
View online: https://doi.org/10.1116/1.5049141
View Table of Contents: https://avs.scitation.org/toc/jva/36/6
Published by the American Vacuum Society

# Responding to the growing issue of research reproducibility

Donald R. Baer[1,a] and Ian S. Gilmore[2,b]
[1]*Pacific Northwest National Laboratory, Richland, Washington 99352*
[2]*National Physical Laboratory, Teddington TW11 0LW, United Kingdom*

(Received 19 July 2018; accepted 28 September 2018; published 12 October 2018)

An increasing number of studies, surveys, and editorials highlight experimental and computational reproducibility and replication issues that appear to pervade most areas of modern science. This perspective examines some of the multiple and complex causes of what has been called a "reproducibility crisis," which can impact materials, interface/(bio)interphase, and vacuum sciences. Reproducibility issues are not new to science, but they are now appearing in new forms requiring innovative solutions. Drivers include the increasingly multidiscipline, multimethod nature of much advanced science, increased complexity of the problems and systems being addressed, and the large amounts and multiple types of experimental and computational data being collected and analyzed in many studies. Sustained efforts are needed to address the causes of reproducibility problems that can hinder the rate of scientific progress and lower public and political regard for science. The initial efforts of the American Vacuum Society to raise awareness of a new generation of reproducibility challenges and provide tools to help address them serve as examples of mitigating actions that can be undertaken. *Published by the AVS.* https://doi.org/10.1116/1.5049141

## I. INTRODUCTION

### A. Is there a reproducibility crisis?

In a survey of 1576 scientists reported in Nature in 2016,[1] 90% of those interviewed indicated they thought there was at least some level of reproducibility problem in the scientific literature. More than half of the scientists responding indicated there was a "significant reproducibility crisis" and only 3% indicated there were no reproducibility problems. In the fall of 2017, National Public Radio's Science Correspondent Richard Harris wrote a guest editorial in Chemical and Engineering News[2] on "Reproducibility issues," observing that although discussion about a "reproducibility crisis" often focused on preclinical medical research and social psychology studies, based on the causes of nonreproducibility there are many indications the issue extends to the physical sciences as well.

Questions or concerns about reproducibility can be asked at multiple levels, but they all address or impact the ability for the results of studies to be replicated by others. Reproducibility issues can generally be categorized into two topics: (1) research integrity, which includes the ethics, misconduct, and deviation from good research practices along with the mechanisms to identify and correct research outputs that fall short of these; and (2) reproducibility or replicability of an experiment, computation, or a whole study. In this discussion, we focus entirely on the latter as the former is already well-developed, for example, the United Kingdom Concordat on Research Integrity[3] and the U.S.A., Federal Research Misconduct Policy.[4]

Uses of the terms reproducibility, repeatability, and replicability have overlapping, distinctive, and inconsistent meanings depending on the area of science.[5] One common distinction is the replicability of a study versus the reproducibility of the analysis of a set of results or the repeatability of a measurement.[6] They all deal in different ways with the reliability and repeatability of scientific results. Issues within the reproducibility and replication category include: (i) Are results within one study appropriately reproducible and appropriately measured and reported, and/or are only selected measurements reported with some others "ignored"? (ii) Have research groups established and communicated adequately clear procedures, protocols, qualified measurement methods, and training to enable others within their group (or close collaborator) to reproduce materials, analysis results, and other aspects of research? (iii) Are reports, measurement approaches, and information in publications adequately detailed so other research teams can reproduce the work?

The issue of reproducibility has received sufficient attention that the U.S. Congress mandated the National Science Foundation to fund a National Academies (NAs) study on *Reproducibility and Replicability in Science*. Their charter is to "explore what is known and identify areas that may need more information to ascertain the extent of reproducibility and replication, review current activities to improve reproducibility and replication highlighting examples of good practices, and examine factors that adversely affect reproducibility and replication." The study group has met four times, and the study webpage contains informative videos of speakers discussing the topic in different fields.[7]

A great deal of discussion and many articles have been dedicated to this topic at a high level and, as noted below, concerns have been noted in specific areas of science. There is a growing recognition that as science and technology have evolved with advanced computation and experimental methods there are new issues and more needs to be done to address them. The term "crisis" is somewhat hyperbolic since research endeavors continue to productively bear fruit in new innovations and products. However, in addition to

[a]Electronic mail: don.baer@pnnl.gov
[b]Electronic mail: ian.gilmore@npl.co.uk

slowing scientific progress, nonreproducibility of high profile and other research results impacts public and political opinion, and the general regard for science that may have long lasting and significant impacts on science and society.

The research community will learn more about the nature and extent of the problem and likely obtain useful guidance from the NAs report and other studies. However, given the systemic nature of the challenges and the possible long-range impacts, it is useful to identify and initiate efforts to improve the current situation. This perspective identifies some of the sources of reproducibility problems and ways the research community can respond now to address them. The perspective is prepared in the context of materials, interface/(bio) interphase, vacuum, and other sciences important to the Journals of Vacuum Science and Technology (JVST) parts A and B, and the American Vacuum Society (AVS) and highlights, as examples, some of the initial actions planned or undertaken in response to reproducibility issues.

### B. Are reproducibility issues relevant to areas of science and technology published in JVST and of importance to the AVS?

As noted in the Harris editorial,[2] the causes of nonreproducibility appear in most areas of science. Important reproducibility challenges in areas of relevance to the JVST and the AVS in both experimental and computational fields were discussed in the NAs sessions including presentations by Kate Kirby and David Sholl during the December 2017 meeting.[8] During his NAs talk on reproducibility, Sholl (a senior editor of Langmuir) noted that publications in materials chemistry and related fields commonly report new materials with limited attention paid to reproducibility. He cited his study of reproducibility of $CO_2$ adsorption on metal organic frameworks where 20% of the data reported could be classified as outliers (significantly different or distant from the majority of other observations).[9] In 2014, the National Institute of Standards and Technology (NIST) highlighted the reproducibility issue in biomedical research in Nature Methods[10] and proposed a route to improve reproducibility was through better measurements, appropriate controls, reference materials, and statistics. Recently, the U.K.'s National Physical Laboratory (NPL) prepared a 2017 Nature editorial on the importance of metrology in the efforts to improve reproducibility.[11] The reproducibility issue has been recognized by editors of AVS' Biointerphases[12] in presenting a list of publications identified as "Biointerphases Tutorials" noting that "The following publications have been identified by the Editors as foundational in furthering understanding of the biointerface. The methods and approaches used for characterization at the biointerface are state-of-the-art and thorough."

In 2018, NIST and NPL co-organized a workshop on "Improving Reproducibility in Research: The Role of Measurement Science."[13] The workshop identified five key recommendations and corresponding actions for National Metrology Institutions in support of the research community. Of particular relevance to the AVS are (i) leadership and participation in intercomparison studies, for example, those enabled by the Versailles Project on Advanced Materials and Standards,[14] that involve industry and academia (the AVS has always been an excellent forum for such studies and the metrology they are based upon), (ii) increased training of young researchers in best measurement practice (the AVS short course, webinars, and other educational programs are effective ways to learn best practices), (iii) the provision of reference data [for example, the AVS Surface Science Spectra (SSS) journal and related eSpectra online tool], and (iv) an increasing importance of adoption of the Findable Accessible Inter-operable & Re-useable (FAIR) principles of data management.[15]

In addition to the general articles and commentaries on nonreproducibility issues, the problem has been highlighted and discussed in domain- or discipline-specific publications, virtual journal issues, and editorials including: industry-academia research,[16] reporting properties of materials and devices,[17] computational research,[18] nano-object production, testing and delivery,[19–22] physics research,[23] and basic and preclinical work.[24]

TABLE I. Systemic drivers of nonreproducibility.

| |
|---|
| **Multidisciplinary and multimethod nature of modern science** |
| • Expertise limitations and/or lack or resources to address all critical areas and methods |
| • Insufficient cooperative/collaborative research |
| **Increased complexity of systems, science questions and tools applied** |
| • Limitations to research design |
| • Need for increasingly large range of analysis tools |
| • Large amounts of data and "black box" data analysis |
| • Publication, peer review, and record taking/reporting limitations |
| • Over reliance on "purchased" supplies without characterization or understanding |
| **High competition for limited resources** |
| • Grant sizes have not increased and are hard to get |
| • Hyper competitive research environment |
| • Low value placed on reproducibility |
| • Wishful thinking |

## II. WHY ARE REPRODUCIBILITY ISSUES IDENTIFIED AS A SIGNIFICANT PROBLEM NOW?

Before looking at possible ways to address the reproducibility issue, it is appropriate to ask about the causes. Why is this a problem now? Are there specific causes that can be addressed? It is useful to recognize that reproducibility challenges are not new to science, but the current form of the issue appears to be systemic and may reflect the increasing complexity of modern science, the ability to collect and process large amounts of data, large scale computation enabled by super computers, and the nature of science

funding at this time.[25] Table I and the summaries below indicate some perceived causes identified in the literature and mentioned during many discussions.

### A. Multidisciplinary nature of current science

Today's scholarship is increasingly multidisciplinary and fast moving, bringing together scientists with widely differing expertise, using different technical languages and techniques. This can lead to experiments planned with inadequate research design[25] and measurements being made without the ability or opportunity to validate them properly.[11]

### B. Increased complexity of science questions and tools

There are increased challenges to both researchers and paper reviewers as a growing number of techniques are required for modern materials and other research. The range of needed skills and expertise can exceed those available in reasonably sized research teams or individual reviewers.[21]

Measurement technology is becoming more powerful and complex. Software and hidden algorithms often stand between the raw data and the user: Numbers are processed and data sets are combined automatically. Tracking and quantifying the uncertainty of the final result can get lost amid all the data crunching. Researchers often treat such tools as a "black box" that generates answers they take on trust and find it harder to have an intuitive feel for when the answers are wrong.[11] This area is associated with the topic of data provenance, which is attracting increased attention.[26]

There are strong concerns about proprietary software with algorithms may not be fully disclosed to the user, and make storage of raw data and reproducible analysis difficult. This issue has been identified by those doing surface analysis, but is also pervasive in most other fields.[27] The FAIR data management principles include not only raw and processed data, but also algorithms, tools, and workflows that led to the resulting processed data.[15]

Issues with data and analysis extend to the publication and review process. For some journals, page limitations limit the inclusion of information and data that might be required for fully describing what is needed to reproduce a data set or study. However, some journals are now able to include unlimited online methods information and provide structured checklists[28] to help eliminate some common confounding issues. Many journals encourage submission of supplemental information that can be accessed online with the published paper. A data-focused journal such as the AVS SSS can be a solution to parts of this issue by providing access to complete data sets that may be only partially available in a standard publication. In an article identifying the seven biggest problems facing science, peer review is identified as a broken process,[25] and there are many ideas on how it can be improved.[29–31] Alternate ways to publish and review science are being explored,[32,33] including publishing in notebook form where it would be possible to "reveal results and methods at the same time, thus effectively publishing the paper and the process that produced them at the same time."[32] Open Science concepts also include open access to data and data analysis processes.[34]

Because of expertise, time, or tool limitations, scientists frequently put too much faith in the ingredients they use. Laboratories often do not run enough controls to identify problems in chemical and other supplies they purchase, including both complex chemicals and common reagents. Both can be problematic.[2,35]

### C. High competition, limited resources, and wishful thinking

Another common driver in science is the hypercompetitive world of academia, with a grave imbalance between the amount of grant money available and the number of laboratories vying for it. If the research evaluation process is flawed, it creates perverse incentives—conscious and unconscious—that can reward "flashy" results over careful substance. The Declaration on Research Assessment was drafted in recognition of the flaws in the current methods of evaluating scientific output. The individuals and institutions signing the declaration seek to improve the ways in which the outputs of scholarly research are evaluated; and they seek to halt the practice of correlating the journal impact factor with the merits of research contribution.[36] One manifestation of the ways research is often evaluated is the dramatic increase in the use of superlatives in scientific abstracts.[2]

In each discipline, the research community, partially through the peer-review process, establishes a culture of what is acceptable. One paper notes that "Materials chemistry and related fields commonly report new materials with limited attention paid to reproducibility."[9]

A root cause is that scientists are human beings, and we tend to see what we want (or expect) to see. As the physicist Richard Feynman noted during a memorable commencement address at the California Institute of Technology in 1974, the scientific method is about finding ways to avoid fooling yourself—"and you are the easiest person to fool."[2]

## III. HOW MIGHT REPRODUCIBILITY ISSUES BE ADDRESSED?

Many people have noted, in different ways and at different times, that significant opportunities often come disguised as major challenges or insolvable problems.[37] What has been described as a reproducibility crisis can provide important opportunities for members of the research community. Based on a large number of discussions, many researchers agree that the issues and challenges associated with reproducibility and replicability in both experimental data and computation are of critical importance to researchers, those who pay for the research, and those who would benefit from the research. The multiple and complex reasons for reproducibility and replication issues that are faced today will significantly impact the future of science, especially younger researchers.

In 2017, the AVS Recommended Practices Committee requested a white paper looking at reproducibility challenges

TABLE II. Actions initiated by the AVS to address reproducibility.

| |
|---|
| **Increase awareness and communicate nature of the challenges** |
| • Newsletter and journal articles |
| • Conduct survey of member perceptions and experience with reproducibility challenges |
| • Subject of 2018 Quantitative Surface Analysis topical conference |
| • Focused Topic for 2019 International Symposium |
| • Topic included in "Lunch and Learn" discussions at international symposia |
| **Tools to address challenge** |
| • Established Subcommittee on Reproducibility as part of Recommended Practices Committee to identify needed actions, including examples of good practices, experimental protocols, and check lists, and access to standards |
| • Initiating a series of tutorial articles for society journals |
| • Participation in standards development and interlaboratory comparison studies |

with the request to explore how the AVS might respond. This perspective has evolved from that white paper along with discussions with academic, national laboratory, and industrial researchers. To be effective, a response to reproducibility issues must recognize the multiple and complex causes and assist researchers, students, research sponsors, instrument vendors, reviewers, and editors in addressing these challenges. Many researchers are involved in defining good practices and documentary standards through the professional society committees and standards organizations such as ASTM Internationa, formerly known as American Society for Testing and Materials, or the International Organization of Standardization (ISO). Use of these practices and standards is an excellent first step to improve measurement repeatability.

The initial efforts of the AVS to address reproducibility problems are summarized in Table II. Awareness and identification of the sources of the problems are essential components of any solution. At the end of his C&E News essay, Richard Harris[2] asks what can a careful scientist do? His answer is "First and foremost, be aware of the conditions around you that may increase the risk of irreproducible results, whether they are bad ingredients, dubious statistical traditions, or outside pressures that can shape behavior." Although they may appear in different and possibly increasingly complex forms, reproducibility issues are not new to science and have given rise to the field of metrology, which must play a significant role in addressing the issue.[11]

The initial AVS responses to the reproducibility crisis have the following objectives:

(i) Increase awareness of the issue and the challenges.
(ii) Expand understanding and communicate the nature and causes of the problem.
(iii) Identify and provide tools for scientists and scientific communities to address the challenges. Such tools might take a variety of forms including: (a) creating examples of vetted good practices and providing easy access to protocols, checklists, and standards; (b) providing training courses for best practice for key techniques; (c) expanding Open Science options beyond Open Access journal articles to include Open Content (archive and make available raw data) and Open Process (show or make available method/analysis used).
(iv) Continue as an active forum for the in-depth discussion of the fundamentals and practicalities of techniques. Maintain or expand participation in intercomparison studies that serve to identify problems and to establish reliable protocols and workflows. Maintain high standards on record keeping and reporting and achieving of data.

The reproducibility issue impacts many activities of the AVS and other professional societies including technical meeting programming and the nature of presentations, journal content and peer review, data records and protocols, and educational activities. Therefore, an AVS response to the "reproducibility crisis" needs to include activities at the annual international symposium, participation in metrology and recommended practices efforts, educational courses and information offered by the Education Committee, and continuing efforts of AVS editors to ask reproducibility questions of authors and reviewers.

Increased automation and machine readability of experimental methods are expected to be important routes to improve reproducibility, and the AVS has an important opportunity to develop ontologies and support open-data formats and open-process information. This might involve existing tools such as SSS/eSpectra, which could provide open-data information. These tools or others might be developed to enable the Open-Process access to algorithms and analysis methods. The JVST is starting to publish tutorial articles that offer expert guidance to new researchers or researchers using new tools. In the coming years, look for events and activities at AVS meetings (for example, see Table II), for papers and other efforts to address this important and challenging issue.

To be effective, the scientific community response to reproducibility issues cannot be one response at one time. An AVS subcommittee on reproducibility issues has been established in the Recommended Practices Committee with efforts to include participation from other committees and groups within the AVS. As the society develops multiple efforts to meet the objectives noted above, we seek your input and ideas (don.baer@pnnl.gov, ian.gilmore@npl.co.uk) on how we can facilitate scientific progress and address the scientific and technological challenges we face.

## ACKNOWLEDGMENTS

The authors acknowledge the inputs and encouragement of the JVST and AVS colleagues, as well as others around the world in identifying the nature and impact of reproducibility issues as well as developing approaches to address them.

[1]M. Baker, Nature **533**, 452 (2016).
[2]R. Harris, Chem. Eng. News **95**, 2 (2017), see https://cen.acs.org/articles/95/i47/Reproducibility-issues.html (accessed 4 October 2018).
[3]*The Concordat to Support Research Integrity* (2012), see https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2012/the-concordat-to-support-research-integrity.pdf (accessed 3 September 2018).
[4]*Federal Research Misconduct Policy|ORI - The Office of Research Integrity* (2018), see https://ori.hhs.gov/federal-research-misconduct-policy (accessed 3 September 2018).
[5]H. E. Plesser, Front. Neuroinform. **11**, 76 (2017).
[6]Replicability Research Group Tel Aviv University, *Replicability vs Reproducibility - Replicability Research Group* (2018), see http://www.replicability.tau.ac.il/index.php/replicability-in-science/replicability-vs-reproducibility.html (accessed 3 September 2018).
[7]H. V. Fineberg, *Reproducibility_and_Replicability_in_Science* (2018), see http://sites.nationalacademies.org/dbasse/bbcss/reproducibility_and_replicability_in_science/index.htm (accessed 3 September 2018).
[8]*Project: Reproducibility and Replicability in Science* (2018), see http://www8.nationalacademies.org/cp/meetingview.aspx?MeetingID=9655&MeetingNo=1&utm_medium=email&utm_source=FYI&dm_i=1ZJN,5CVQA,OXY0HK,KPF8H,1 (accessed 3 September 2018).
[9]J. Park, J. D. Howe, and D. S. Sholl, Chem. Mater. **29**, 10487 (2017).
[10]A. L. Plant, L. E. Locascio, W. E. May, and P. D. Gallagher, Nat. Methods **11**, 895 (2014).
[11]M. Sené, I. Gilmore, and J.-T. Janssen, Nature **547**, 397 (2017).
[12]A. Belu, K. Maniura, and S. McArthur, Biointerphases **11**, 040201 (2016).
[13]RDAlliance, *Improving Reproducibility in Research: The Role of Measurement Science* (2018), see https://www.rd-alliance.org/improving-reproducibility-research-role-measurement-science (accessed 3 September 2018).
[14]*Technical Working Area 2 Surface Chemical Analysis* (2018), see http://www.vamas.org/twa2/index.html (accessed 3 September 2018).
[15]M. D. Wilkinson *et al.*, Sci. Data **3**, 160018 (2016).
[16]B. R. Jasny *et al.*, Science **357**, 759 (2017).
[17]J. M. Buriak, C. W. Jones, P. V. Kamat, K. S. Schanze, G. C. Schatz, G. D. Scholes, and P. S. Weiss, Chem. Mater. **28**, 3525 (2016).
[18]R. D. Peng, Science **334**, 1226 (2011).
[19]A. E. Nel *et al.*, ACS Nano **9**, 5627 (2015).
[20]D. R. Baer *et al.*, Surf. Interface Anal. **40**, 529 (2008).
[21]D. R. Baer *et al.*, J. Vac. Sci. Technol. A **31**, 050820 (2013).
[22]D. R. Baer, P. Munusamy, and B. D. Thrall, Biointerphases **11**, 04B401 (2016).
[23]B. Knuteson, Phys. World **30**, 17 (2017).
[24]C. G. Begley and J. P. A. Ioannidis, Circ. Res. **116**, 116 (2015).
[25]B. Belluz, B. Plumer, and B. Resnick, *The 7 Biggest Problems Facing Science, According to 270 Scientists* (2016), see https://www.vox.com/2016/7/14/12016710/science-challeges-research-funding-peer-review-process (accessed 3 September 2018).
[26]D. Durbin and N. Car, *Research Data Provenance IG* (2018), see https://rd-alliance.org/groups/research-data-provenance.html (accessed 3 September 2018).
[27]S. Wild, "Irreproducible astronomy," Phys. Today (2018, April), see https://physicstoday.scitation.org/do/10.1063/PT.6.1.20180404a/full/ (accessed 3 September 2018).
[28]Checklists work to improve science, Nature **556**, 273 (2018).
[29]J. D. Linton, Res. Policy **45**, 1936 (2016).
[30]M. Bionaz, J. Nutr. Food Sci. **3**, 1000e1120 (2013).
[31]R. Harris, *Scientists Aim to Pull Peer Review Out of the 17th Century* (2018), see https://www.npr.org/sections/health-shots/2018/02/24/586184355/scientists-aim-to-pull-peer-review-out-of-the-17th-century
[32]J. Somers, *The Scientific Paper is Obsolete* (2018, April), see https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/ (accessed 3 September 2018).
[33]R. Kiley, *Wellcome Open Research* (2018), see https://wellcomeopenresearch.org/ (accessed 3 September 2018).
[34]G. B. de al Fuente, *What is Open Science? Introduction* (2018), Foster Facilitate Open Science Training for European Research, see https://www.fosteropenscience.eu/content/what-open-science-introduction (accessed 3 September 2018).
[35]R. M. Crist, J. H. Grossman, A. K. Patri, S. T. Stern, M. A. Dobrovolskaia, P. P. Adiseshaiah, J. D. Clogston, and S. E. McNeil, Integr. Biol. **5**, 66 (2013).
[36]San Francisco Declaration on Research Assessment (DORA) (2018), see https://sfdora.org/ (accessed 3 September 2018).
[37]*John W. Gardner Quotes* (2018), see https://www.brainyquote.com/quotes/john_w_gardner_134413 (accessed 3 September 2018).