Reliability of Scientific Knowledge

This section examines the importance of reliability of scientific knowledge, for the scientific community and for society. Reproducibility is one criterion for reliability of scientific knowledge, but reproducibility and reliability are often conflated. In some research fields, reliability does not, and cannot, rely solely on reproducibility. In addition, reproducibility has different meanings that need to be clearly identified. There are three main causes of a lack of reproducibility: the inherent variability of the object being studied, experimental variability, and poorly defined methods.

Research Reproducibility

Marcus Munafò
Professor of Biological Psychology, University of Bristol

Watch this short presentation
to support your work on this topic >>

Reproducibility

Reproducibility—the extent to which consistent results are observed when scientific studies are repeated—is one of science's defining features, and has even been described as the "demarcation criterion between science and nonscience". In principle, the entire body of scientific evidence could be reproduced independently by researchers following the original methods and drawing from insights gleaned by prior investigators. In this sense, belief in scientific evidence is not contingent on trust in its originators. Other types of belief depend on the authority and motivations of the source; beliefs in science do not.

Considering its central importance, one might expect replication to be a prominent part of scientific practice. It is not. An important reason for this is that scientists have strong incentives to introduce new ideas but weak incentives to confirm the validity of old ideas. Innovative findings produce rewards of publication, employment, and tenure; replicated findings produce a shrug.

Open Science Collaboration (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science*, 7(6), 657 – 60.

Re-thinking reproducibility

I take issue with the widespread reference to reproducibility as an overarching epistemic value for science and a good proxy measure for the quality and reliability of research results. Reproducibility comes in a variety of forms geared to different methods and goals in science ... variation is linked to the degree of control that researchers are able and willing to exercise on their materials and on the environment in which their investigations take place, as well as to the extent to which they rely on statistical methods in assessing the validity of the evidence being produced. In studies where control over environmental variants is only partially achieved, for instance, reproduction resulting in different outcomes is perceived as highly valuable, since it can signal hitherto unknown sources of variation or define the scope of the hypothesis being tested.

By contrast, in studies that are carried out in highly idiosyncratic environmental conditions and/or on perishable and rare samples which do not lend themselves to statistical analysis, it is the very uniqueness and irreproducibility of research conditions that makes the resulting data valuable as sources of evidence. In such cases, a focus on enhancing reproducibility turns out not to be the best way to foster high-quality, robust research outcomes. Rather, it is the well-informed analysis of how reliable and meaningful data are obtained through irreproducible research practices that increases the sophistication of research methods and of the ways in which they are documented and disseminated.

Leonelli, S. (2018). Re-Thinking Reproducibility as a Criterion for Research Quality. Forthcoming in *Research in the History of Economic Thought and Methodology, Special Issue "Curiosity, Imagination and Surprise".*

PERFORM researcher reflection

66 **I found it interesting to consider the increasing issue of reproducibility in science. It highlights the issues within the academic system, where what is good for science is not necessarily good for scientists. As a scientist I feel that awareness of the issue is key, so that changes and planning can be put in place to try and increase reproducibility as much as possible. Reproducibility in my field of work can and has been an issue...** 99

Activities

Before starting discussion, ask everyone to pick one of the research projects they are working on and summarise it in 2-3 sentences. You may choose to focus on one of the discussion topics below, or you may want to address each of them in turn.

Understanding reliability

1. Write on post-it notes: In your research project, according to which criteria do you consider whether the research knowledge that you produce is reliable? (e.g. reproducibility of the results, cross-checking of the methodology by peers, consistency of the results with other studies, clear understanding of the possible biases, etc.). Write one criterion per post-it. Spread the post-its on a board.
2. Have a collective discussion on the following questions:

   - Is there a diversity of reliability criteria within your research group, for instance depending on the methods that are used?
   - What are the reliability criteria that are specific to your research field? Is there a person in charge of research integrity in your institute?
   - When do you doubt the reliability of your own results, or of the results presented by colleagues?

Understanding reproducibility

1. Read in advance or collectively read 'Re-Thinking Reproducibility as a Criterion for Research Quality' quoted above.
2. Invite everyone to write their definition of reproducibility on a piece of paper (e.g. ability to obtain the same results, ability to replicate the experiment, ability to replicate the analysis, etc.). Share everyone's definitions. The facilitator can introduce the publication 'What does research reproducibility mean?' referenced below, at this point.
3. Have a collective discussion on the following questions:

   - Is reproducibility a relevant criterion for the reliability of your research? Does your answer change depending on different interpretations of 'reproducibility'?
   - In which scientific disciplines do you consider reproducibility an important criterion for reliability? Does your answer change depending on different interpretations of 'reproducibility'?
   - (If appropriate:) In your field, what are the possible explanations if, when replicating an experiment or a study, you don't obtain the same results?

Reliability: why does it matter and for whom?

1. Carry out a quick survey on the following question and invite participants to provide precise examples to support their answer: Do you consider that the research published in your field is very/ reasonably/ moderately/ poorly reliable?
2. Discuss the following questions, asking participants to present examples to illustrate their opinion:

   - Why does it matter for the scientific community that results are reliable?
   - Why does it matter for society that scientific results are reliable?
   - How should scientists communicate with people about the reliability of scientific results?
   - What are the risks with scientists overplaying the reliability of their results?

References and additional resources

- Baker, M. (2016). *Is there a reproducibility crisis? Nature, 533(7604), 452-454*. DOI: 10.1038/533452a
- Goodman, S. N., Fanelli, D., & Ioannidis, J.P.A. (2016). What does research reproducibility mean? *Science Translational Medicine, 8(341), 341ps12*. DOI: 10.1126/scitranslmed.aaf5027
- Leonelli, S. (2018). Re-Thinking Reproducibility as a Criterion for Research Quality. Forthcoming in *Research in the History of Economic Thought and Methodology, Special Issue "Curiosity, Imagination and Surprise"* https://goo.gl/omtT8r
- Makri, A. (2017). Give the public the tools to trust scientists. *Nature, 541(7637)*, 261. DOI:10.1038/541261a
- Open Science Collaboration (2012). *An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. Perspectives on Psychological Science, 7(6)*, 657 – 60. DOI:10.1177/1745691612462588