

Reproducibility made simple

Automating reproducible research workflows

Aaron Peikert

2020-03-11

Contents

Abstract	5
1 Introduction	7

Abstract

This is a short summary.

Chapter 1

Introduction

Claerbout & Karrenbach define reproducibility as the ability to gain the same results, from the same dataset. In contrast, they call a result replicable if one draws the same conclusion from a new dataset. This thesis concerns itself with the former, providing researchers with an accessible analysis workflow, that is virtually guaranteed to reproduce across time and devices. The scientific community agrees that their work should be ideally reproducible. Indeed it may be hard to find a researcher who distrusts a result because it is reproducible; to the contrary, many feel it is “good scientific practice” to ensure it is. Several reasons, practical and meta-scientific, justify this consensus of reproducibility as a minimal standard of Science.

Reproducibility makes researchers life more productive in two ways: The act of reproduction provides, at the most basic level, an opportunity to spot errors, helping the researchers who originally produced them. At the same time, other researchers may benefit from reusing materials from an analysis they reproduced.

Beyond these two purely pragmatic reasons, reproduction is crucial, depending on the philosophical view of Science one subscribes to, because it allows independent validation and enables replication. Philosophers of Science characterise Science by a shared method of determining if a statement about the world is “true” or more broadly evaluating the statements verisimilitude. If this method is for experts to agree on the assumptions and deduce some truth, reproducibility is hardly necessary. On the other hand, it gains importance if one induces facts by carefully observing the world. The decisive difference is that the former gains credibility through the authority of the experts, while the latter is trustworthy because anyone may verify it. That Science should provide facts general enough to be theoretically verifiable by anyone is an argument deeply persuasive to me. Some have even argued that this democratisation of Science is what fueled the scientific revolution. The scientific revolution had the experiment

as an agreed-upon method to observe the reality and a much later revolution provides statistical modelling as a means to induction. This consensus, about how to observe and how to induce, gives modern scientific enterprises much of its credibility. These two reasons, justify why we must assume reproducibility as a scientific standard if we accept induction as a scientific method: First, it enables independent verification of the process of induction, and secondly, it dramatically simplifies replication as a means to verify the induced truths.

However, neither the practical reasons that results may be less error-prone and more reusable nor the meta-scientific grounds that the process of induction and the induced facts are more straightforward to verify, if reproducible, follow strictly from the definition given above. Imagine a binary program that is perfectly reproducible; hence upon input of the same dataset, it fills a scientific manuscript with the same numbers at the right places. Furthermore, assume this hypothetical program may never hold if the data changes. Does the predicate “reproducible” here reduce the number of mistakes or enables reuse? Unlikely. Or could one audit it and use it in replication? Hardly. This admittedly constructed case shows us: we are not interested in reproducibility, we are interested in its side effects.

Spoiling its elegant simplicity, I change the definition by Claerbout & Karrenbach to address this issue, by further demanding that reproducibility must facilitate replication. Hence, I would call a result only then reproducible if the results remain unchanged if the data does, and it furthermore helps other researchers to replicate the results if they attempt to do so. With such a notion, the only valid cause of reproducibility is transparency. Because only if it is clear how the data relates to its results, both reproducibility and replication get promoted. It follows that something is no longer either reproducible or not, but there are shades, because a research product may promote replication to varying degrees. Note, that a scientific result can facilitate replication without anyone ever attempting to replicate it, e.g. by educating other researches about the analyses method, being openly accessible and providing reusable components. This much more demanding standard of reproducibility may gain justification by two recent developments in the social sciences in general and psychology in particular: the emergence of a “replication crises” and the rise of “machine learning” as a scientific tool. Both trends link to the use of statistical modelling on which the social sciences became reliant for testing and developing their theories. It turns out, if one fits the same statistical model as published on newly gathered data, one fails to achieve the same results as published more often than not. Such failure to replicate findings previously believed to be robust has amounted to a level some social scientists call a crisis. They put forth various causes and remedies to this crisis. Most remedies share a common theme: transparency. Some call for Bayesian statistics, as it makes assumptions more explicit, or demand preregistration as a means to clarify how to analyse the data, beforehand and publicly, others require the researchers to publish their data. These calls for transparency, as a response to the replication crises, formed the open science movement which stresses the necessity of six principles:

- Open Access
- Open Data
- Open Source
- Open Methodology
- Open Peer Review
- Open Educational Resources

I argue that a research product resting on these pillars facilitates replication the most and hence satisfies the highest standard of reproducibility. If everyone has access to a scientific product and its data along with the source code, leading them to understand the methodology and thus enabling them to criticise the result and educate themselves, one is in the best position to replicate it. Hence, any one's ability to reproduce such result gives a tangible affirmation of its usefulness to the scientific community.

Reproducibility is nothing special when anyone can perform the calculations needed with a pocket calculator; however, the more and more frequent use of computer-intensive methods renders such expectation questionable. The use of machine learning techniques, once enabled by the computer taking over strenuous works, now impedes our quest for reproducibility. More massive amounts of more complicated computer code than ever create room for errors and misunderstandings. Yet, I am far from calling for abstinence from machine learning, just because it complicates reproduction, but want to emphasise the need for solutions that allow anyone to reproduce even the most sophisticated analysis.

Peikert & Brandmaier (2020) put forth an analysis workflow which provides just this accessibility for everyone to reproduce any analysis. However, they fail to provide the same level of convenience for the researcher who created an analysis in the first place. Setting up the workflow eats up a considerable chunk of the researchers time, which they may better spend at advancing research. This additional effort offsets the increasing productivity, promised by reproducibility, which I regard as most significant in the workflows adoption. Persuading researchers, who find the meta-scientific argumentation noble but impractical, do not care about it or oppose it, requires concrete, practical benefits. Luckily, most of this setup process may be automated, letting the researcher enjoy the workflows advantages while decreasing the efforts necessary to achieve them. Providing an easier to use and more accessible version of the analysis workflow by Peikert & Brandmaier (2020) is the goal of this thesis and the herein presented repro package for the R programming language.