Reproducibility by Other Means: Transparent Research Objects

Timothy M. McPhillips, Craig Willis, Michael R. Gryk,
Santiago Nuñez-Corrales, and **Bertram Ludäscher**

Workshop on Research Objects 2019 (RO2019)
IEEE eScience 2019, San Diego CA
24 September 2019

Is reproducibility really so complicated?

- Reproducibility crisis?
- Terminology crisis?
- Or gullibility crisis?
- What is reproducibility anyway?
- And who is responsible for it?

What is the single most effective way to make your research more reproducible?

a) Carefully record and report your work.

b) Use open source software exclusively and make any new or modified code freely available.

c) Employ the latest interoperability standards for scientific data, metadata, software, and Research Objects.

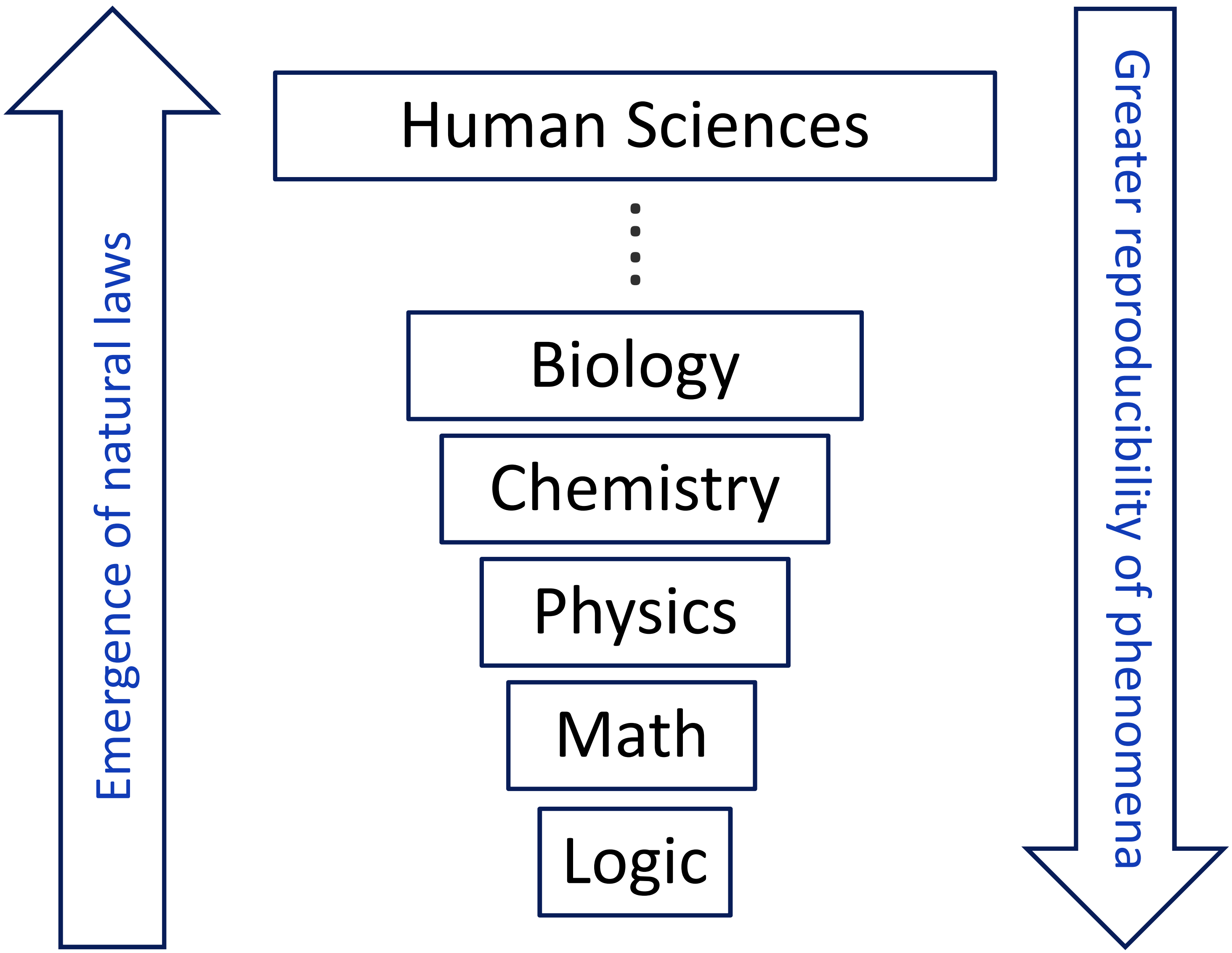
d) Do all of your work in software containers.

e) Focus your research on intrinsically reproducible phenomena.

Basic assumptions made by researchers in the natural sciences

- We are discovering things that are the way they are whether we go and look for them or not.
- We are discovering things that conceivably could be different than they happen to be. To find out how things actually are we must go look.
- It does not matter who does the looking. Everyone with the same opportunity to look will find the same things to be true.

A (snobbish) hierarchy of intrinsic reproducibility?



It’s not so simple…

Limits on reproducibility in the natural sciences

- Nature is not a digital computer. It's more of an entropy generator built on chaos and (true) randomness with natural laws, math, and logic serving as constraints.
- Good experiments are hard to design and to perform even once.
- Instruments can be costly and limited in supply.
- Many phenomena cannot be studied via experiment at all.
- Past events are crucial to many theories.
- Some things happen only once.

What is always possible? Transparency.

FASEB* definition of transparency

Transparency: *The reporting of experimental materials and methods in a manner that provides enough information for others to independently assess and/or reproduce experimental findings.*

- Transparency is what allows an experiment to be reviewed and assessed independently by others.
- Transparency *facilitates* reproduction of results but does not *require* reproduction to support review and assessment.
- It is considered a *problem* if exact repetition of the steps in reported research is required either to evaluate the work or to reproduce results.

* The Federation of American Societies for Experimental Biology comprises 30 scientific societies and over 130,000 researchers.

Quantifying repeatability

- Experiments on natural phenomena generally are *not* exactly repeatable.
- Materials, conditions, equipment, and instruments all vary.
- Uncertainty is intrinsic to most measurements.
- Experimental biologists perform **replicate experiments** to assess end-to-end repeatability.

  Technical replicates: Measurements and data analyses performed on the **same sample** using the **same equipment** multiple times.

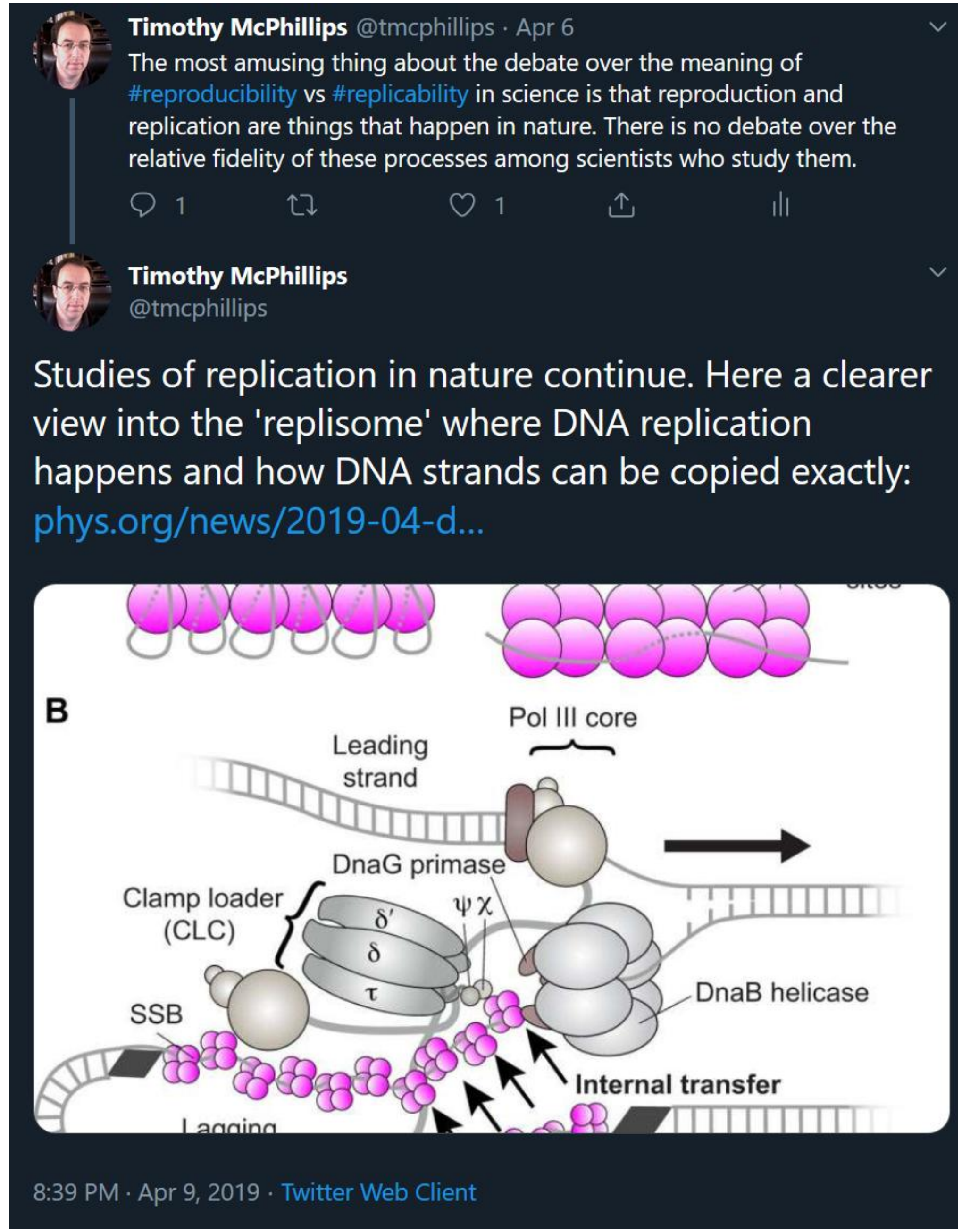
  Biological replicates: Measurements and data analyses performed on **different** but **biologically equivalent samples** on the **same equipment**.

Why are these “replicates”, not “reproductions”?

Replication and reproduction are natural processes that biologists study

- Probably the most amazing aspect of life is the incredible fidelity with which genetic material—DNA—is *replicated* within cells.
- DNA replication is carried out by the *replisome*—which even detects and corrects errors on the fly!
- Organisms *reproduce* and have *reproductive* systems.
- Biological reproduction is *much* lower fidelity than DNA replication. In fact, the process of reproduction often *encourages variation* in the children.

Experimental **replicates** assess the **highest possible fidelity** at which an experiment can be **repeated**—by the same researcher, using the same equipment, on the same or equivalent samples, immediately one after the other in time.



Theorists talk about replication

- Dawkins' ***selfish genes*** are *replicators*.
- Debate in **origins of life** research:

  Did replication or metabolism come first?
- Could life have started before high-fidelity replication of genetic material was achieved?
- For these theorists and philosophers *high-fidelity* is the defining characteristic of *replication*.

Stanford Encyclopedia of Philosophy

Replication and Reproduction

First published Wed Dec 5, 2001; substantive revision Tue Sep 25, 2018

The problem of replication and reproduction arises out of the history of genetics [see the entry gene for a historical review]. It is tied to the concept of the gene and its generalization in an evolutionary context [see the entry evolution]. Richard Dawkins introduced the notion of replicators—things that self-replicate—as a universalization of evolutionary understandings of genes. Dawkins argued that replicators are the *sine qua non* of evolution by natural selection [see the entry natural selection], while other accounts only require *reproduction* as one of its defining features. What exactly is a replicator? How are replicators different from genes? Can evolution by natural selection occur without the existence of replicators? Besides the biological domain, are there any other domains in which replicators have been postulated? To answer these questions, we will first provide some background for Dawkins' notion of replicator and its ties with the concepts of the gene and information. We will then introduce the distinction between *Replicators* and *Vehicles* in the context of biological evolution and followed by the extension of this to other domains. Finally, we will discuss some of the challenges to the idea that replicators are necessary conditions for evolution by natural selection.

FASEB* definitions of reproducibility and replicability

Maximal fidelity to original **experiment**, greater fidelity to original result.

Replicability: *The ability to* ***duplicate*** *(i.e., repeat) a* ***prior result*** *using the* ***same source materials*** *and* ***methodologies****. This term should only be used when referring to repeating the results of a* ***specific experiment*** *rather than an entire study.*

Reproducibility: *The ability to achieve* ***similar*** *or nearly identical* ***results*** *using* ***comparable materials*** *and* ***methodologies****. This term may be used when specific* ***findings from a study*** *are obtained by an independent group of researchers.*

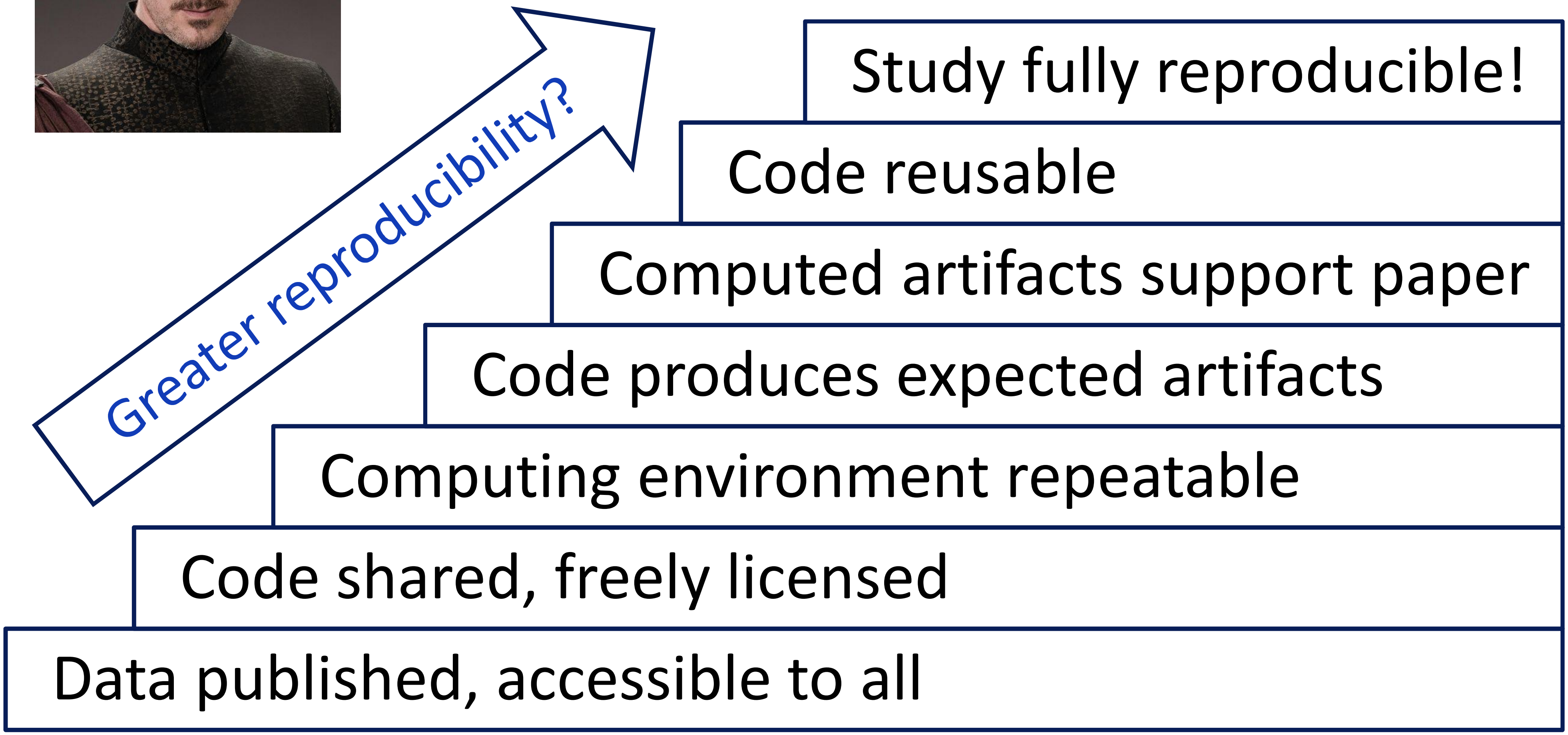
Less fidelity to original **study**, lower fidelity result expected.

***** The Federation of American Societies for Experimental Biology comprises 30 scientific societies and over 130,000 researchers.

Beyond reproduction and replication: exact repeatability

- Digital computers use logic gates to achieve replication of information at such a low error rate we can call it *exact*.
- Computers pull the exactness of logic and discrete mathematics up to the level of macroscale phenomena– quite a feat.
- Exactness is (effectively) achievable for computer hardware, compiled software, program executions, and computing environments.
- Researchers employing digital computers have access to a new kind of reproducibility never before seen in science: *exact repeatability*.

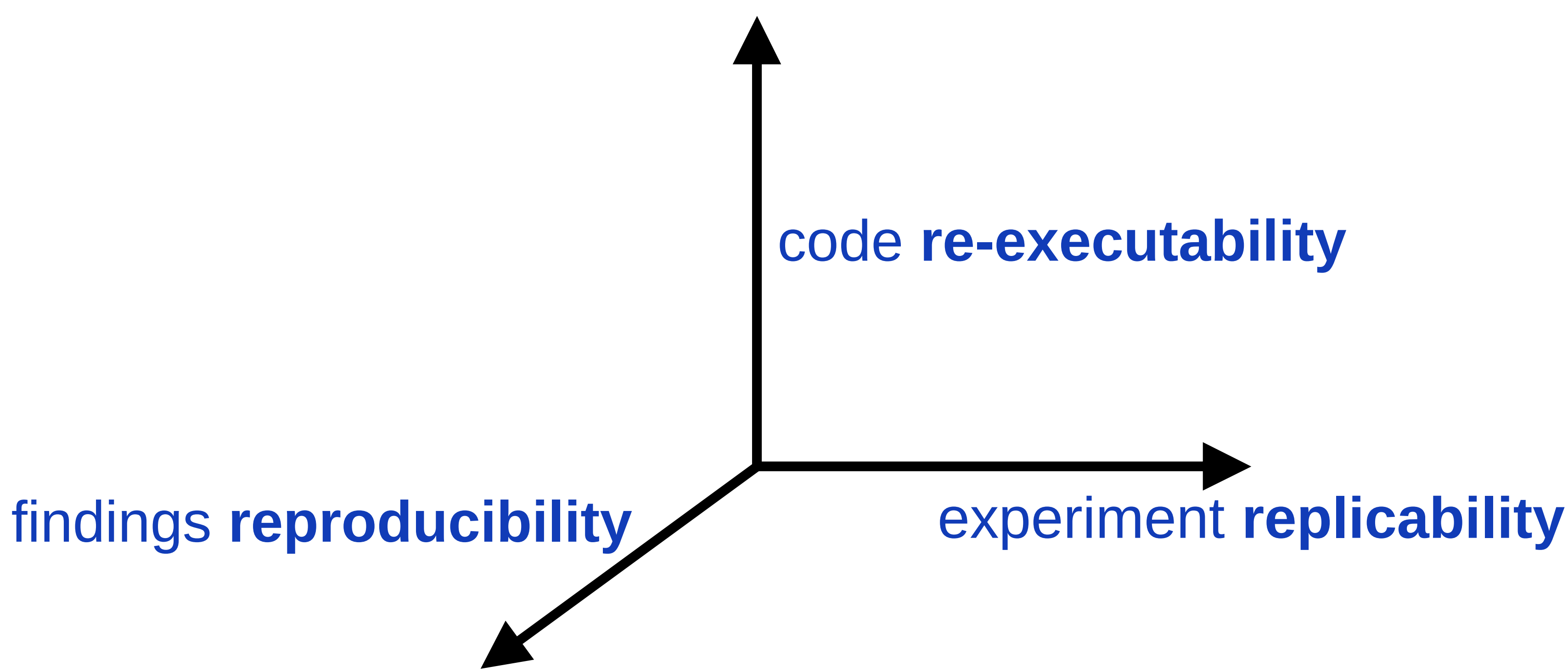
Chaos is a ladder. Is reproducibility a staircase?



It is tempting to think about reproducibility one-dimensionally.

But what if scientific reproducibility is multidimensional?

- Do the R-words have an obvious order, where achieving one must precede achieving the next?
- Or might they represent basis vectors of some kind of multidimensional space?



National Academy of Sciences definitions of reproducibility and replicability

Maximal fidelity to original **computation**, greater fidelity to original result.

Reproducibility *is obtaining* ***consistent results*** *using the* ***same input data****,* ***computational steps****,* ***methods****, and* ***code****, and* ***conditions*** *of analysis.*

Replicability *is obtaining* ***consistent results across studies*** *aimed at answering the* ***same scientific question****, each of which has obtained its* ***own data****.*

Less fidelity to original **study**, different data.

These definitions:

- **Reverse relative fidelity** of *reproducibility* and *replicability* compared to FASEB definitions. Replicability is lower fidelity. Headache for biologists.
- **Require code** from reproducibility. NAS report explicitly **equates** *reproducibility* and *computational reproducibility*.
- Leave **non-computational research** components with only **one word**, *replicability*, analogous to *FASEB::reproducibility*.
- Provide **no** way of expressing biologists' concept of **experimental replicates**—without a computer.

Modeling reproducibility as multidimensional may offer way out of the terminology quagmire

- Recognize that **different terminologies refer to different sets of dimensions**; communities focus on different subspaces, or different choices of basis vectors.
- **Map conflicting definitions onto shared dimensions**; use mappings to convert claims made using one terminology to claims using a different terminology.
- Allow each community to focus on dimensions of interest to them using the most intuitive terminology; **use namespaces to eliminate ambiguity**.
- Use Research Objects to **attach claims about reproducibility to research artifacts**, to **disambiguate** these **claims**, and to **support queries** using terminology of the user's choosing.

Reproducibility badges and verification workflows – yet more variation

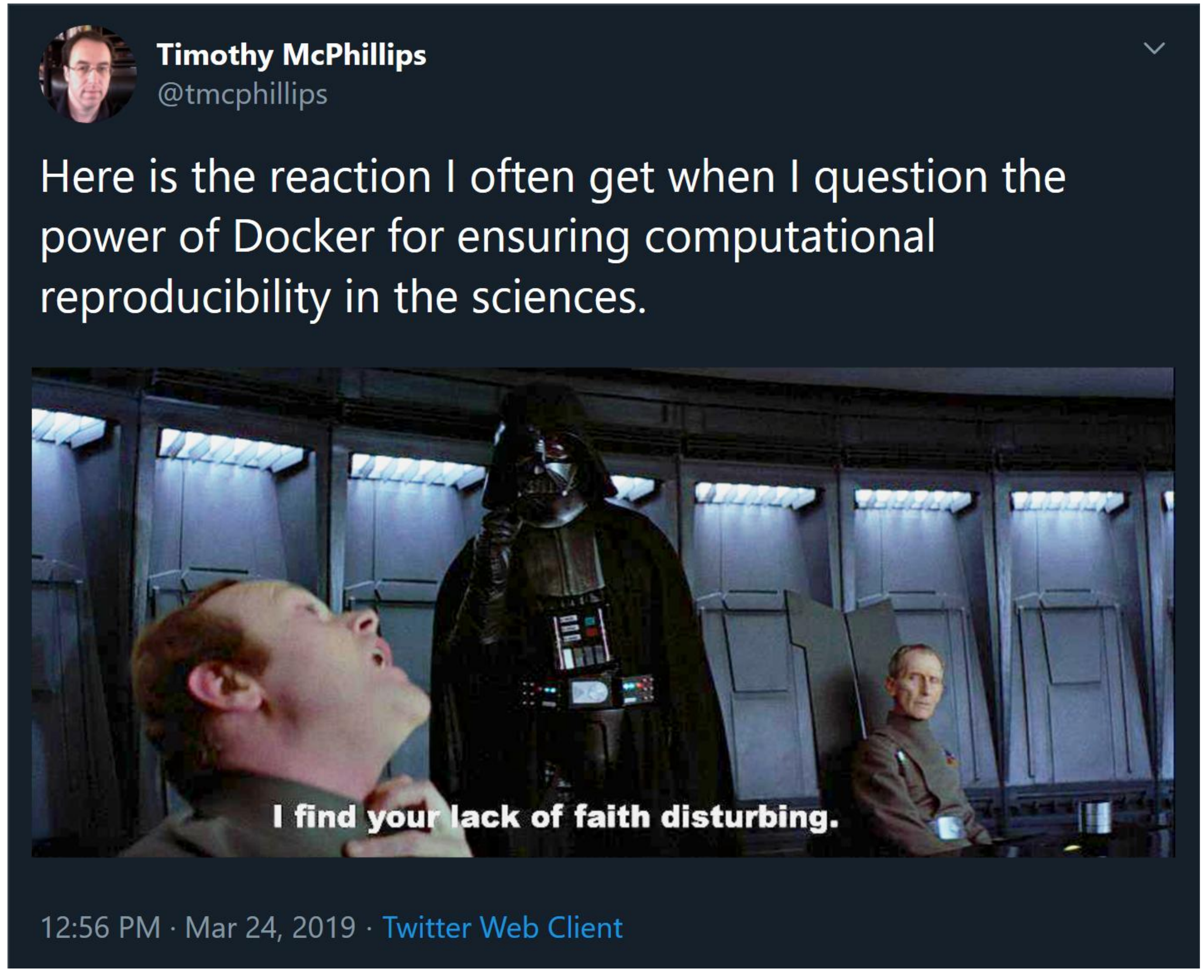
- ACM **SIGMOD** defines a defines a procedure for **assessing** database **research reproducibility**.
- **ACM awards four different reproducibility badges** distinct from the SIGMOD reproducibility assessment.
- **ACM** has defined **8 versions** of the guidelines for awarding its badges **since 2015**.
- The workflow used by the American Journal of Political Science (**AJPS**) to **verify computational artifacts** also is **versioned**.
- The **meanings** of reproducibility badges **change from year to year** even within a single organization—with no end in sight.

If we want these badges to have any meaning at all they must be mapped to something that isn't constantly changing.

db-reproducibility.seas.harvard.edu, www.acm.org/publications/policies/artifact-review-badging, ajps.org/wp-content/uploads/2019/01/ajps-quant-data-checklist-ver-1-2.pdf

Computational reproducibility claims often are ambiguous

- **Current approaches** for preserving computing environments **may not work for long**.
- A **Dockerfile** that builds correctly today **might not** do so a year from now–if it builds at all.
- **Transitive dependencies** on 3rd-party shared libraries lead to particularly **fragile** software **builds**–even if you pin the versions of your direct dependencies.
- So **what do we really mean** when we say we have made computing environments, software, or computational products “reproducible”?



We need to map terminologies for computational reproducibility onto dimensions that will outlive particular technologies.

Transparent Research Objects

- Transparency in the natural sciences **enables research to be evaluated**—and reported results used with confidence—**without actually repeating others' work**.
- How can Research Objects **extend the advantages of transparency to computational research** and the computational components of experimental studies?
- Researchers need to be able to **query** the **reproducibility characteristics** of artifacts in Research Objects.
- These queries need to be poseable **using terminology familiar to the researcher**—terminology likely different from that used by the author of the Research Object.
- Queries about computational reproducibility need to take the **longevity of technological approaches to reproducibility** into account.