Reproducible Research: Pfff... Why bother?

Arnaud Legrand
CNRS, Inria, University of Grenoble

March 4th, 2018 – POLARIS days, Allevard

Common beliefs

- *RR mainly allows to fight scientific misconduct (e.g., fraud). That's nice but I'm honnest so just let me do my work!*

Common beliefs

- *RR mainly allows to fight scientific misconduct (e.g., fraud). That's nice but I'm honnest so just let me do my work!*
- *RR is all about re-executing the same code, even if the code is stupid and makes wrong computation. It's pointless!*

Common beliefs

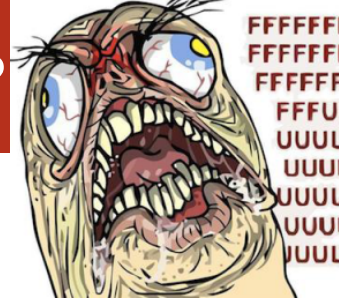
- *RR mainly allows to fight scientific misconduct (e.g., fraud). That's nice but I'm honnest so just let me do my work!*
- *RR is all about re-executing the same code, even if the code is stupid and makes wrong computation. It's pointless!*
- *My student has done everything with org-jupyter-studio-mode. Now he's gone and I can't reuse what he did. See, what's the point? RR does not help!*

Common beliefs

- *RR mainly allows to fight scientific misconduct (e.g., fraud). That's nice but I'm honnest so just let me do my work!*
- *RR is all about re-executing the same code, even if the code is stupid and makes wrong computation. It's pointless!*
- *My student has done everything with org-jupyter-studio-mode. Now he's gone and I can't reuse what he did. See, what's the point? RR does not help!*
- *RR is about controling and checking everything, which slows down the scientific discovery process. Changing the way we work and publish may be harmful!*

Why did I get interested in this in the first place ?

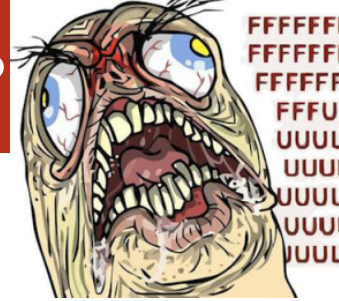
Why did I get interested in this in the first place ?



Frustration as an author, a reviewer, a scientist

- The damned fourth reviewer asked for a major revision and wants me to change figure 3. ☹ Which code and which data set did I use to generate this figure?
- 6 months later: Why did I do that?
- There is no label/legend/... What is the meaning of this graph? If only I could access the generation script and get rid of the logscale
- This average value must hide something. As usual, no confidence interval... I wonder whether the difference is significant at all
- How does this really work ? It this improvement solely the result of this naive idea?

Why did I get interested in this in the first place ?



Frustration as an author, a reviewer, a scientist

- The damned fourth reviewer asked for a major revision and wants me to change figure 3. ☹ Which code and which data set did I use to generate this figure?
- 6 months later: Why did I do that?
- There is no label/legend/... What is the meaning of this graph? If only I could access the generation script and get rid of the logscale
- This average value must hide something. As usual, no confidence interval... I wonder whether the difference is significant at all
- How does this really work ? It this improvement solely the result of this naive idea?

Is your Frustration really a good motivation for annoying everyone else?

A recent story

February 2018: *could you have a look at this article for a review?*

The Pitfalls of Provisioning Exascale Networks: A Trace Replay Analysis for Understanding Communication Performance

UQ and HPC simulation! Exactly what I'm interested in at the moment!

A recent story

February 2018: *could you have a look at this article for a review?*

The Pitfalls of Provisioning Exascale Networks: A Trace Replay Analysis for Understanding Communication Performance

UQ and HPC simulation! Exactly what I'm interested in at the moment!

Wait, I obviously know who wrote this. **Flashback:**

- June 2013: Validation and Uncertainty Assessment of Extreme-Scale HPC Simulation through Bayesian Inference (EuroPar)
- October 2016: Dagstuhl seminar on UQ and HPC. Explanations from Habib Najm, a UQ expert! ☺

A recent story

February 2018: *could you have a look at this article for a review?*

The Pitfalls of Provisioning Exascale Networks: A Trace Replay Analysis for Understanding Communication Performance

UQ and HPC simulation! Exactly what I'm interested in at the moment!

Wait, I obviously know who wrote this. **Flashback:**

- June 2013: Validation and Uncertainty Assessment of Extreme-Scale HPC Simulation through Bayesian Inference (EuroPar)
- October 2016: Dagstuhl seminar on UQ and HPC. Explanations from Habib Najm, a UQ expert! ☺

Back to 2018:

- Same vague (but convincing!) explanations, still no code, just as useless to me as 4.5 years ago
- There is no scientific misconduct at all. Yet, it is useless.
  - none of the other reviewer complained about this
  - why would the authors bother? (*researchware*)

(For me) RR = **sharing** research results

Wait, I thought this is what conferences and articles were about! :)

The five R's of RR: Re-Run, Repeat, Reproduce, Replicate, Reuse

Many technical aspects:

- software/environment engineering, preservation, and continuous integration
- numerical aspects
- experimental aspect (measure, statistics)
- provenance tracking and information/code/data sharing

In my case: allow informed inspection

The laboratory notebook and a better understanding of statistics are essential

What can I do to change all this ?

- SMPE lecture (CS master students)
- Keynotes (mostly toward computer scientists, PhD, postdocs, . . .)
  - even at the Inria Scientific Days in 2016
- Webinars
- Program committees in conferences
  - Should I sign the PRO manifesto ? https://opennessinitiative.org/
  - RepPar but now many "competing" workshops
- Inria strategical plan: 1 out of 21 is on RR
- MOOC: in June on FUN ?
  - a much larger target audience, hence teach the basics
  - journaling, literate programming, simple provenance tracking/backup
- Keynote for DGD-T Inria (SED and STIP engineers) in May 2018
- HRS4R: 1st meeting in March 2018

I'm glad to do this but this is exhausting 

Does RR qualify for scientific research by the way ?

Knowing about all these tools definitely allows me to improve my research. Is this enough ?

- RR is mentioned in the POLARIS proposal as a transverse activity, not as a research axis.
- Epistemic opacity: is it a scientific question or a social problem?
- Will it be solved all by itself ? Are we really doing RR ourselves ? Do we care ?

Any thoughts ?