



# Next in Reproducibility: standards, policies, infrastructure, and human factors

 @LorenaABarba

# About me

- ▶ Reproducibility PI Manifesto  
*figshare*, 2012
- ▶ “The hard road to reproducibility”  
*Science*, Oct. 2016
- ▶ “Repro Packs”  
*Nature blogs*, Apr. 2017
- ▶ SC19 Reproducibility Chair
- ▶ NASEM Committee member

Lorena A. Barba group

**Reproducibility PI Manifesto**



<http://lorenabarba.com>



CONSENSUS STUDY REPORT

# Reproducibility and Replicability in Science



- ▶ Study mandated by public law 114-329 (Jan. 2017)
- ▶ commissioned by the National Science Foundation (NSF) to The National Academies of Sciences, Engineering and Medicine (NASEM)
- ▶ 15 experts convened
- ▶ 18 months of in-person meetings, teleconferences, commissioned papers, deliberations, writing
- ▶ report released 7 May 2019

<http://doi.org/c5jp>



# **Defining Reproducibility & Replicability**

# Def.— Reproducibility

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







## **Jon F. Claerbout**

*Professor Emeritus of Geophysics  
Stanford University*

... pioneered the use of computers in processing and filtering seismic exploration data [Wikipedia]

... from 1991, he required theses to conform to a standard of reproducibility.



# “In 1990, we set this sequence of goals:

1. Learn how to **merge a publication with its underlying computational analysis**.
2. Teach researchers how to prepare a document in a form where they themselves can reproduce their own research results a year or more later by “pressing a single button”.
3. Learn how to leave finished work in a condition where coworkers can reproduce the calculation including the final illustration by pressing a button in its caption.
4. Prepare a complete copy of our local software environment so that graduating students can take their work away with them to other sites, press a button, and reproduce their Stanford work.

“I’ve learned that interactive programs are [tyranny] (unless they include the ability to arrive in any previous state by means of a script).”

— Jon Claerbout





# Def.— Replicability

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

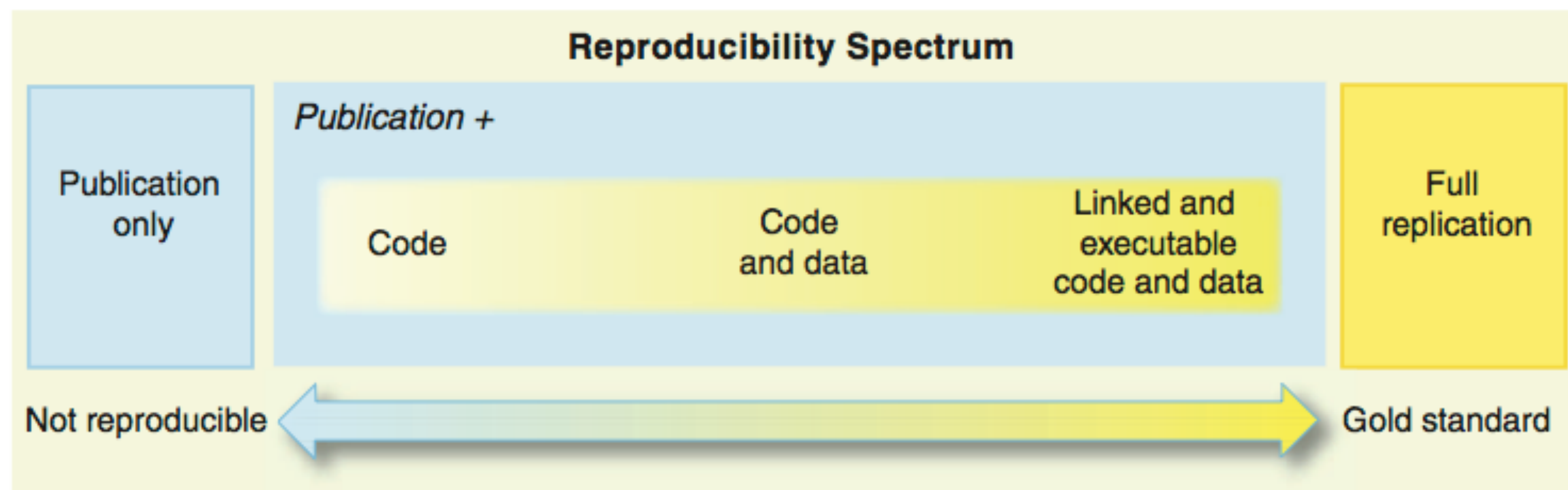




PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng



**Fig. 1.** The spectrum of reproducibility.



# **Reproducibility**



# Widespread use of computation & data in science



- As important as the telescopes were the software libraries and data products needed to create the first image of a black hole

(now iconic photo of Dr. Katie Bouman)



- ▶ 92% of academics use research software
- ▶ 69% say that their research would not be practical without it
- ▶ 56% develop their own software
- ▶ 21% of those have no training in software development

S.J. Hettrick et al. (2014), *UK Research Software Survey* doi:10.5281/zenodo.14809

“reproducibility . . . requires having the complete software environment [...] and the full source code available for inspection, modification, and application under varied parameter settings.”

—Buckheit and Donoho (1995)



RECOMMENDATION 4-1: To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis, unless such information is restricted by non-public data policies. That information should include the data, study methods, and computational environment.

FINDING 4-4: Understanding the limits of computational reproducibility in increasingly complex computational systems, such as artificial intelligence, high- performance computing, and deep learning, is an active area of research.



RECOMMENDATION 4-2: The National Science Foundation should consider investing in research that explores the limits of computational reproducibility in instances in which bitwise reproducibility is not reasonable in order to ensure that the meaning of consistent computational results remains in step with the development of new computational hardware, tools, and methods.

# Sources of non-reproducibility

- ▶ Inadequate record keeping
- ▶ Nontransparent reporting
- ▶ Obsolescence of the digital artifacts
- ▶ Flawed attempts to reproduce other's results
- ▶ Barriers in culture



# Improving reproducibility

- ▶ Automatic capture of computational details; workflow management systems
- ▶ Source code and data version control
- ▶ Tools for reproducing results via virtualization, cloud computing, packaging, containers (e.g., Docker, Singularity)
- ▶ Interactive computational notebooks (e.g., Jupyter)

RECOMMENDATION 6-3: Funding agencies and organizations should consider investing in research and development of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion. Concurrently, investments would be helpful in outreach to inform and train researchers on best practices and how to use these tools.





A set of open-source tools  
for **interactive** and  
exploratory computing.

# Jupyter grant proposal:

“...the core problem we are trying to solve is the collaborative creation of **reproducible** computational narratives.”





**Interactive** → ← **Reproducible**



Interactive → ← Reproducible





**Philip Stark**

@philipbstark Follows you



**Philip Stark**

@philipbstark

Following

Relying on Excel for important calculations is like driving drunk: no matter how carefully you do it, a wreck is likely. [#reproducibility](#)

1:14 AM - 11 Aug 2014

41 Retweets 38 Likes



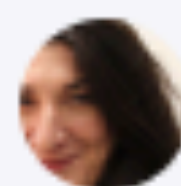
4



41



38



Tweet your reply



**Philip Stark** @philipbstark · 11 Aug 2014

Replying to [@philipbstark](#)

2\

*On spreadsheets:*

**“...the user interface conflates input, output, code, and presentation, making testing code and discovering bugs difficult.”**

— Philip Stark, *Science is ‘show me,’ not ‘trust me’* (2015)



Why do we care about  
Computational Reproducibility?

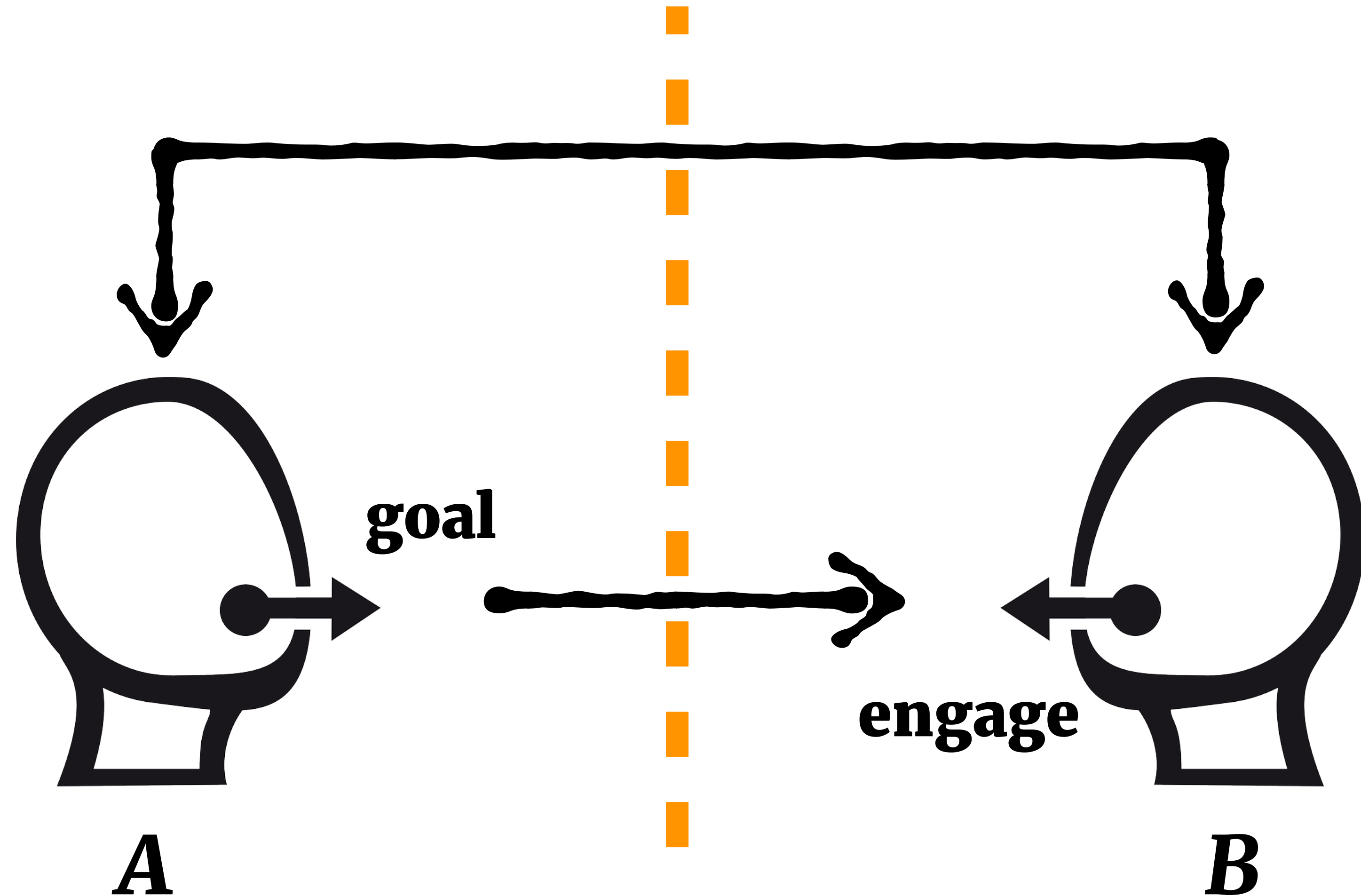
# **“Science is a conversation”**

—Stephen Downes (“connectivism”)

- ▶ a conversation between scientists and their body of knowledge
- ▶ a conversation among scientists
- ▶ a conversation between scientists and machines...



# What is a conversation?



How do we design  
(conversations) for  
reproducibility?

Herbert A. Simon

**The Science of Design:  
Creating the Artificial**





**Ben Shneiderman**



# “Designing the User Interface”

—Ben Shneiderman, 6th ed.

Tools that succeed are:

- ▶ comprehensible,
- ▶ predictable, and
- ▶ controllable

Those who have authority and responsibility must have adequate levels of control.

*Responsibility* should guide design.

**Human control ↓ ↑ Automation**

“Ensuring human control while increasing automation.”

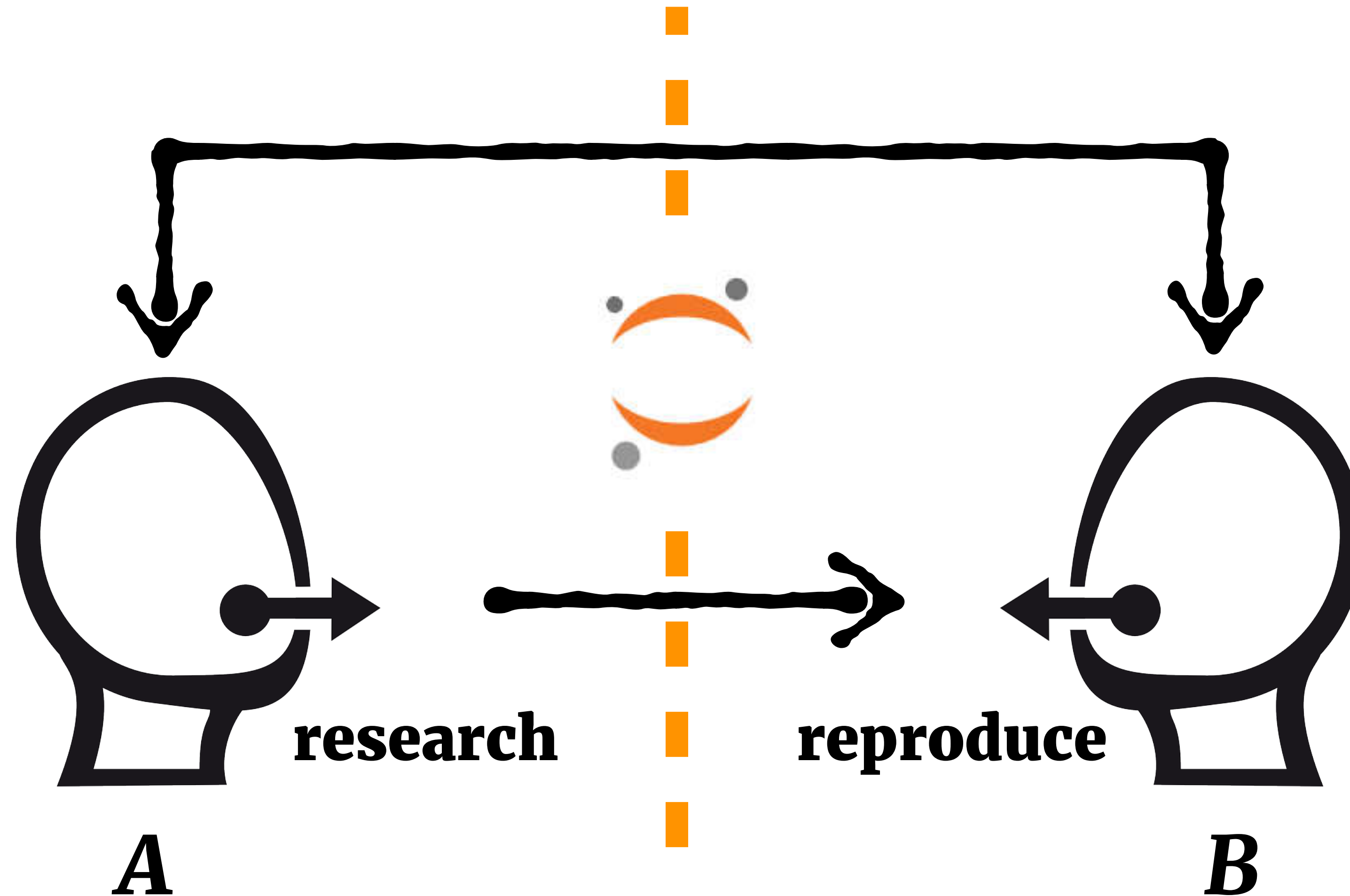
*On 21st-century design:*

**“...design has expanded from giving form to creating systems that support human interactions.”**

— Hugh Dubberly & Paul Pangaro,  
*Cybernetics and Design: Conversations for action* (2015)



# Conversation builds *trust*



*“I have a button here. I push the button.  
That’s not a conversation.”*

— Paul Pangaro,  
*Rethinking Design Thinking, PICNIC Festival  
Amsterdam (2010)*



**Reproducibility:**  
**not a one-click solution**



# “Algorithmic Accountability”

—Ben Shneiderman, 2017 Turing Lecture

Independent oversight (for complex projects):

- ▶ planning oversight (e.g., building)
- ▶ continuous monitoring (e.g., banks)
- ▶ retrospective analysis (e.g. NTSB)

# Human Factors

# In aviation

- ▶ Human factors key to safety
- ▶ Built into system design and operations management
- ▶ Imperfect human performance is the root cause in the majority of aviation accidents
- ▶ human-factors training place focus on improving cooperation among workers
- ▶ important to develop a “safety culture,” including standard of reporting and open sharing of data on incident and solutions



# Design considerations

- ▶ cockpit automation
- ▶ flight path management systems
- ▶ recovery from failure

# Behavior

- ▶ confirmation bias
- ▶ complacency
- ▶ lack of communication
- ▶ distractions
- ▶ human error types

# Organization

- ▶ safety culture
- ▶ performance pressures
- ▶ technical competencies (regular assessment and training)





# Next in Reproducibility: standards, policies, infrastructure, and human factors

 @LorenaABarba