Table of Contents

Table of Contents

Motivation: Prevent Loss of Knowledge

Every semester, graduate students around **the world** take an Empirical/Applied [...] Economics course. A typical assignment consists of reproducing the results of a paper and, possibly, testing the robustness of its results.

Stage	New Knowledge
Scope (select and verify)	Data and code exist?
Assess	Degree of reproducibility for specific part of the paper
Improve	E.g. fixed paths, libraries, added missing files, etc.
Test robustness	Results are robust to additional specifications

Table of Contents

Context for ACRE

- American Economics Association (AEA) creates first data policy in 2006.

  - Must publish some data (waivers available)

- AEA updates policy in 2019.

  - Must post all data and code. Publication is conditional on verifying reproducibility (if confidential: must document extensively)
  - A new requirement is to post all cleaning code, even for data that is not public
  - See the AEA Data Editor Website for more information

- We should expect high levels of computational reproducibility after 2019 (AEA Journals).

- We should not demand 100% reproducibility before, but we could identify the gaps and try to improve some.

Beyond Binary Judgments

Reproductions can easily gravitate towards adversarial exchanges.

- Early career researcher (ECR) have incentives to emphasize unsuccessful reproductions
- Original authors have a more senior position and can use it to deter in-depth reproductions from ECRs.
- The media also focuses on eye-catching headlines

Our approach:

We do not want to say

> "Paper X is (ir)reproducible"

We do want to say

> "Paper X's result Y has a high/low **level** of reproducibility according to **several** reproduction attempts. Moreover, **improvements** have been made to the original reproduction package, **increasing** its reproducibility to a higher level"
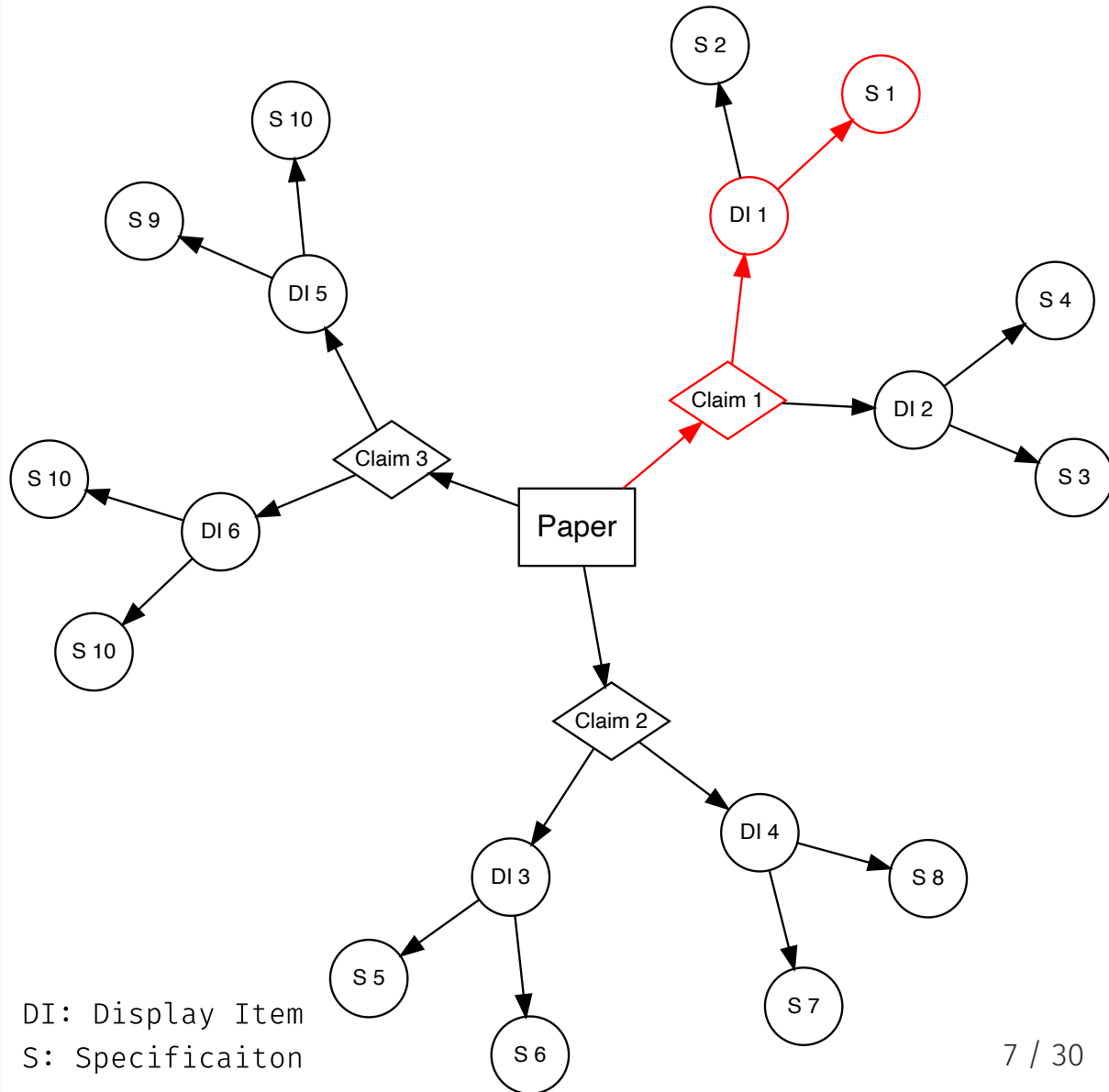
Our Framework

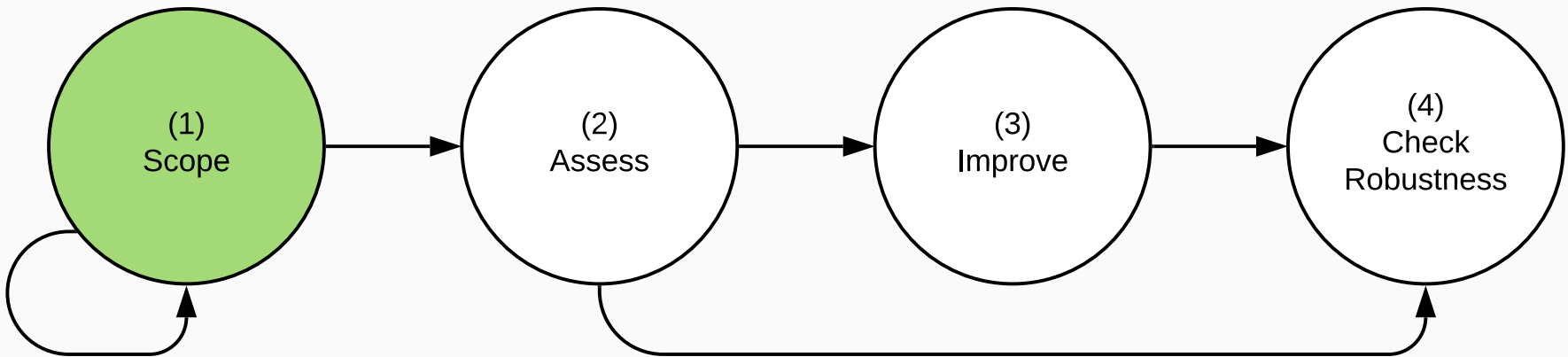Each **reproduction attempt** is centered around scientific **claims**

One paper can contain several claims.

Each claim may be supported by various **display items**: tables, figures & inline results.

A reproduction attempt is at the claim level, and reproducers must record their **specifications** of interest.

DI: Display Item
S: Specificaiton

Stages

Scoping

1. Select or be assigned a candidate paper

2. Check ACRE Platform for previous entries and verify availability of reproduction package (RP)

3. If no RP, leave a short record, and repeat with a different candidate paper

4. Once RP is found then candidate becomes declared paper

5. Only then: read the paper and select claim(s), display items and specification to reproduce
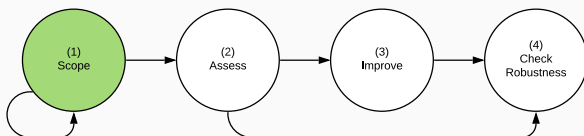
Box 1: Summary Report Card for ACRE Paper Entry
Title: Sample Title
Authors: Jane Doe & John Doe
Original Reproduction Package Available: URL/No
[If "Yes"]
Additional Reproduction Packages: Number (eg., 2)
[If "No"]
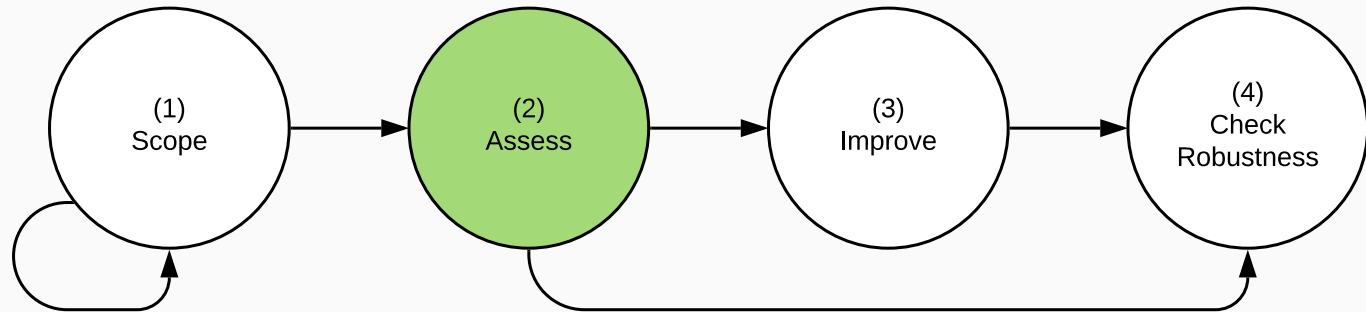**Contacted Authors?:** Yes/No
[If "Yes(contacted)"]
Type of Response: Categories (6).
Authors Available for Further Questions for ACRE Reproductions: Yes/No/Unknown

Assessment

Two main parts for assessment:

1. Find all the elements behind a display item
2. Score the reproducibility of that display item

Identify All the Elements Behind a Display Item

Reproducers will be asked to draw a clear connection to the raw data sources mentioned in the paper and the display item under reproduction.

Data sources

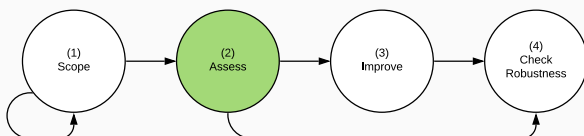Connect the data sources in the paper's text with specific raw data files.

Analytic data sets

Describe each analytic data file.

Code files

Inspect all code files and record all their inputs and outputs.

With all the information recorded above, reproducers can use the **ACRE Diagram Builder** to generate a **reproduction tree**.

Reproduction Tree

```
table1.tex
    |___[code] analysis.R
        |___analysis_data.dta
            |___[code] final_merge.do
                |___cleaned_1_2.dta
                |    |___[code] clean_merged_1_2.do
                |        |___merged_1_2.dta
                |            |___[code] merge_1_2.do
                |                |___cleaned_1.dta
                |                |    |___[code] clean_raw_1.py
                |                |        |___raw_1.dta
                |                |___cleaned_2.dta
                |                    |___[code] clean_raw_2.py
                |                        |___raw_2.dta
                |___cleaned_3_4.dta
                    |___[code] clean_merged_3_4.do
                        |___merged_3_4.dta
                            |___[code] merge_3_4.do
                                |___cleaned_3.dta
                                |    |___[code] clean_raw_3.py
                                |        |___raw_3.dta
                                |___cleaned_4.dta
                                    |___[code] clean_raw_4.py
                                        |___raw_4.dta
```
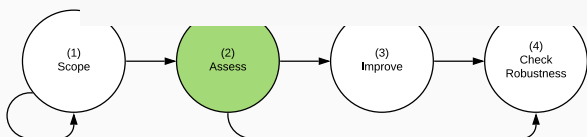
Levels

```
              Levels of Computational Reproducibility
              (P denotes "partial", C denotes "complete")
```

Level	Analysis Code (P / C)		Analysis Data (P / C)		CRA	Cleaning Code (P / C)		Raw Data (P / C)		CRR
L1: No materials	–	–	–	–	–	–	–	–	–	–
L2: Only code	✔	✔	–	–	–	–	–	–	–	–
L3: Partial analysis data & code	✔	✔	✔	–	–	–	–	–	–	–
L4: All analysis data & code	✔	✔	✔	✔	–	–	–	–	–	–
L5: Reproducible from analysis	✔	✔	✔	✔	✔	–	–	–	–	–
L6: Some cleaning code	✔	✔	✔	✔	✔	✔	–	–	–	–
L7: All cleaning code	✔	✔	✔	✔	✔	✔	✔	–	–	–
L8: Some raw data	✔	✔	✔	✔	✔	✔	✔	✔	–	–
L9: All raw data	✔	✔	✔	✔	✔	✔	✔	✔	✔	–
L10: Reproducible from raw data	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

(1) Scope → (2) Assess → (3) Improve → (4) Check Robustness
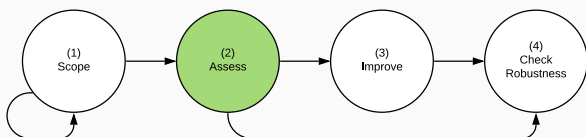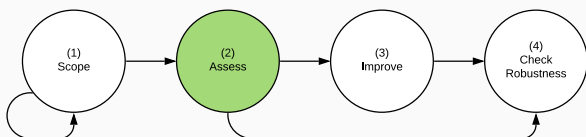
```
                  Levels of Computational Reproducibility
                     with Proprietary/Confidential Data
                   (P denotes "partial", C denotes "complete")
                            | Availability of materials, and reproducibility |
                            |------------------------------------------------|
                            |         | Instr.  |     |          | Instr.|    |
                            |Analysis| Analysis|     | Cleaning| Raw   |    |
                            |Code    | Data    | CRA | Code    | Data  | CRR|
                            | P | C  | P | C  |     | P | C  | P | C |    |
                            ---------|---------|-----|---------|-------|-----|
        L1: No materials.................| -   -  | -   -  | -  | -   - | -   - | -  |
        --------------------------------|--------|---------|-----|---------|-------|-----|
        L2: Only code ...................| ✔   ✔  | -   -  | -  | -   - | -   - | -  |
        L3*: Partial analysis data & code| ✔   ✔  | ✔   -  | -  | -   - | -   - | -  |
        L4*: All analysis data & code....| ✔   ✔  | ✔   ✔  | -  | -   - | -   - | -  |
        L5*: Proof of third party CRA....| ✔   ✔  | ✔   ✔  | ✔  | -   - | -   - | -  |
        --------------------------------|--------|---------|-----|---------|-------|-----|
        L6: Some cleaning code...........| ✔   ✔  | ✔   ✔  | ✔  | ✔   - | -   - | -  |
        L7: All cleaning code............| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔ | -   - | -  |
        L8*: Some instr. for raw data....| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔ | ✔   - | -  |
        L9*: All instr. for raw data.....| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔ | ✔   ✔ | -  |
        L10*:Proof of third party CRR....| ✔   ✔  | ✔   ✔  | ✔  | ✔   ✔ | ✔   ✔ | ✔  |
```
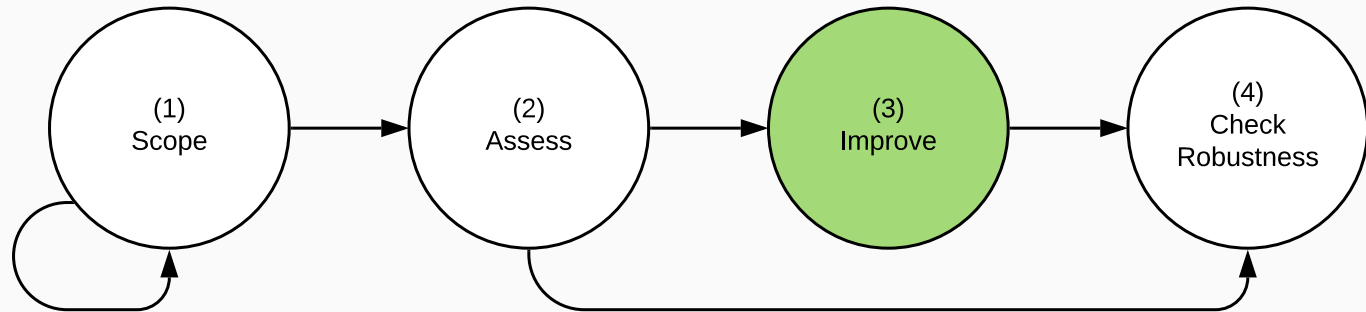
(1) Scope → (2) Assess → (3) Improve → (4) Check Robustness
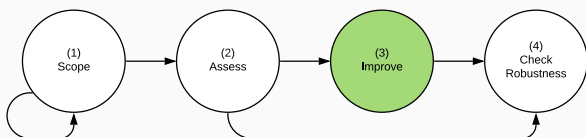
Improvements

Three types of improvements:

1. Improvements at the paper level
2. Improvements at the display-item level
3. Specific future improvements

Improvements: Paper-level

- Use version control software (Git/Github).
- Improve documentation: comments, indentations, object names, etc.
- Re-organize the reproduction package into a set of folders and sub-folders that follow standardized best practices, and add a master script that executes all the code in order, with no further modifications. See AEA's reproduction template.
- Literate programming environment (e.g., Jupyter notebooks, RMarkdown)
- Re-write code using a differenet statistical software (ideally open source, like R, Python, or Julia).
- Set up a computing capsule (e.g., Binder and Code Ocean).

(1)
Scope

(2)
Assess

(3)
Improve

(4)
Check
Robustness

Improvements: Display item-level

- Adding missing raw data: files or meta-data
  - Example: "Add raw temperature and relative humidity data"
- Adding missing analytic data files
  - Example: "Copy the row files from Data folder into new `Analysis\trade cost\Input`"
- Adding missing analysis or cleaning code
  - Example: "Replaced broken Wald bootstrap code with updated code/command"

- Debugging code

  - Example: "was counting each group 4 times in round 1, so fixed that"

- Propose specific future improvements
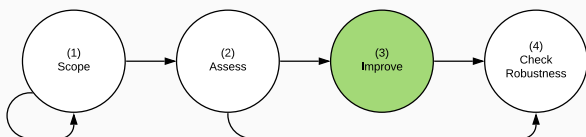
Robustness Checks
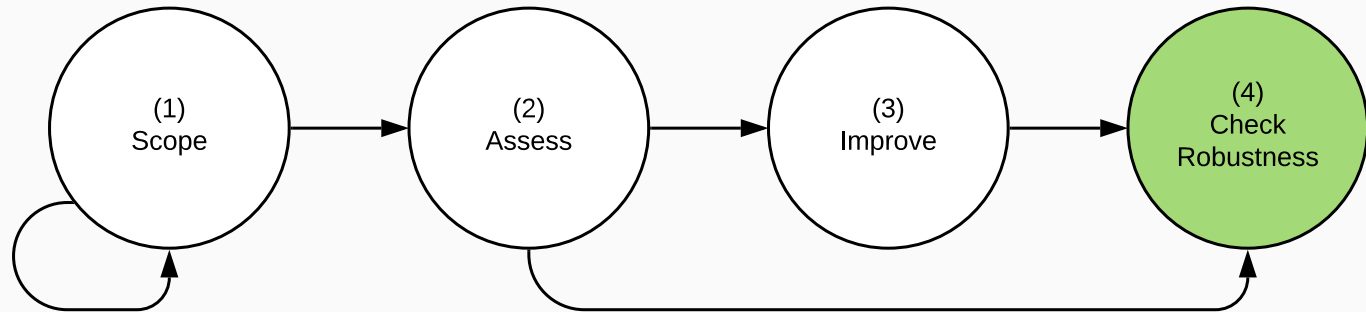
Two main parts for robustness:

1. Increase the number of possible robustness checks
2. Justify the appropriateness of a specific test

Robustness

Robustness checks: any possible change in a computational choice, both in data analysis and data cleaning

Reasonable specifications (Simonsohn et. al., 2018):

1. Sensible tests of the research question
2. Expected to be statistically valid, and
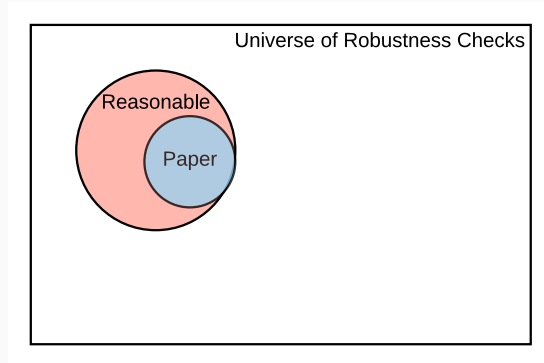3. Not redundant with other specifications in the set.

Reproducers will be able to record two types of contributions:

- Mapping the universe of robustness checks
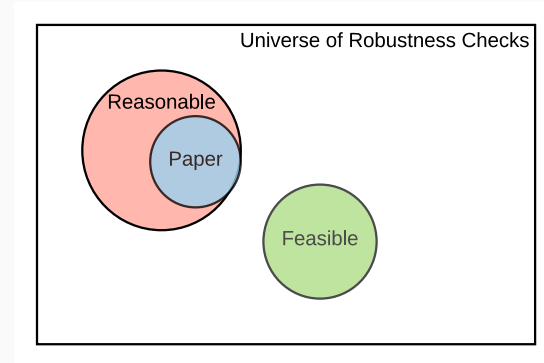- Justify a specific robustness check as reasonable
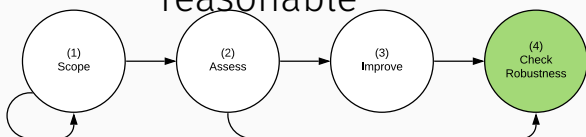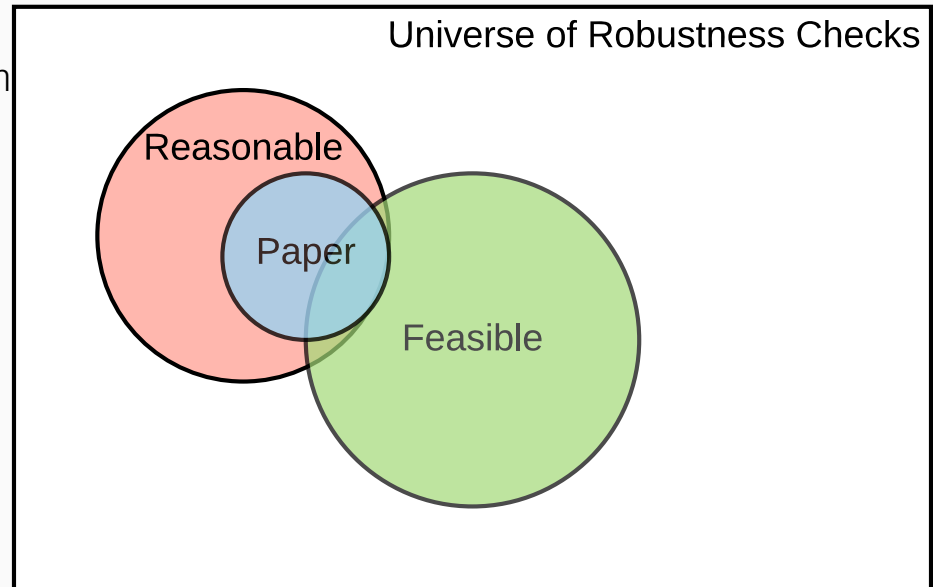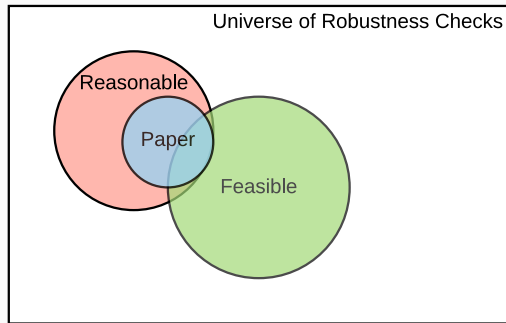
Robustness & Reproducibility
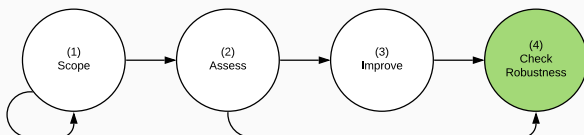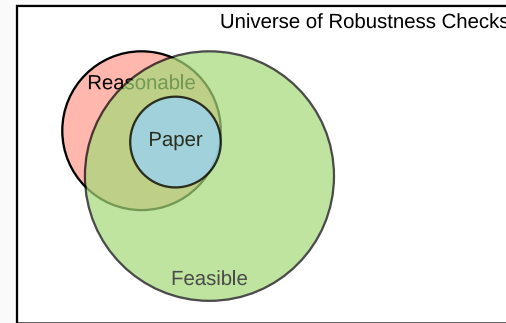
Robustness with level 1

Robustness with levels 2-4

Robustness with levels 5-9

Robustness with level 10

Promoting a Constructive Exchange

1 - Contacting the original author(s) when there is no reproduction package

2 - Contacting the original author(s) to request specific missing items of a reproduction package

3 - Asking for additional guidance when some materials have been shared

4 - Response when the original author has refused to share due to *undisclosed reasons*

5 - Response when the original author has refused to share due to legal or ethical restrictions of the data

6 - Contacting the original author to share the results of your reproduction exercise

7 - Responding to hostile responses from original authors

Under development: sample responses form authors to reproducers

Example 1: There is no reproduction package

Subject: Reproduction package for [`"Title of the paper"`]

Dear Dr. [`Lastname of Corresponding Author`],

I am contacting you to request a reproduction package for your paper titled [`Title`] which was published in [`Reference`]. A reproduction package may contain (raw and/or analytic) data, code, and other documentation that makes it possible to reproduce paper. Would you be able to share any of these items?

I am a [`position`] at [`Institution`], and I would like to reproduce the results, tables, and other figures using the reproduction materials mentioned above. I have chosen this paper because [`add context ... `]. **Unfortunately, I was not able to locate any of these materials on the journal website, Dataverse** [`or other data and code repositories`], **or in your website**.

I will record the result of my reproduction attempt on ACRE [...]. With your permission, I will also record the materials you share with me, which would allow access for other reproducers and avoid repeated requests directed to you. Please let me know if there are any legal or ethical restrictions that apply to all or parts of the reproduction materials so that I can take that into consideration during this exercise.

In addition to your response above, would you be available to respond to future (non-repetitive) inquiries from me or other reproducers conducting an ACRE excercise? **Though your cooperation with my and/or any future request would be extremely helpful, please note that you are *not required to comply*.**

Since I am required to complete this project by [`date`], I would appreciate your response by [`deadline`].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [`Name (email)`] for further details on this exercise. Thank you in advance for your help!

Best regards,
[`Reproducer`]

Template email:

> **Subject:** Clarification for reproduction materials for [“Title of the paper”]
>
> Dear Dr. [Lastname of Corresponding Author],
>
> Thank you for sharing the materials. They have been immensely helpful for my work.
>
> Unfortunately, I ran into a few issues as I delved into the reproduction exercise, and I think your guidance would be helpful in resolving them. **[Describe the issues and how you have tried to resolve them. Describe whatever files or parts of the data or code are missing. Refer to examples 1 and 2 below for more details]**.
>
> Thank you in advance for your help.
>
> Best regards,
> [Reproducer]

An example of well described issues:

> Specifically, I am attempting to reproduce [`display item X (e.g., table 1, figure 3)`]. I found that the following components are required to reproduce to reproduce [`display item X`]:

```
display_item_X
    └──[code] formatting_table1.R
       ├──output1_part1.txt
       │    └──[code] output_table1.do
       │       └──[data] analysis_data01.csv
       │          └──[code] data_cleaning01.R*
       │             └──[data] UNKNOWN
       └──output1_part2.txt
            └──[code] output_table2.do
               └──[data] analysis_data02.csv
                  └──[code] data_cleaning02.R
                     └──[data] admin_01raw.csv*
```

I have marked with an asterisk (*) the items that I could not find in the reproduction materials: **data_cleaning01.R** and **admin_01raw.csv**. After accessing these files, I will also be able to identify the name of the raw data set required to obtain output1_part1.txt. This is to let you know that I may need to contact you again if I cannot find this file (labeled as **UNKNOWN** above) in the reproduction materials.

I understand that this request will require some work for you or somebody in your research group, but I want to assure you that I will add these missing files to the reproduction package for your paper on the ACRE platform. **Doing this will ensure that you will not be asked twice for the same missing file.**

Ok, I get it. But what is in for me?

- Standardized homework/project: everything is set up in terms of structure and deliverables.

- Easy to grade (homework format).

- Easy to guide and oversee (undergraduate dissertation format).

- Easy to setup as an independent study.

- Reduces duplication of requests to authors.

- Facilitates a constructive exchange of ideas.

  - When emailing authors.
  - When discussion reproduction attempts.

- Personal satisfaction that you're contributing a public good to the profession!

Easy to grade: report 1

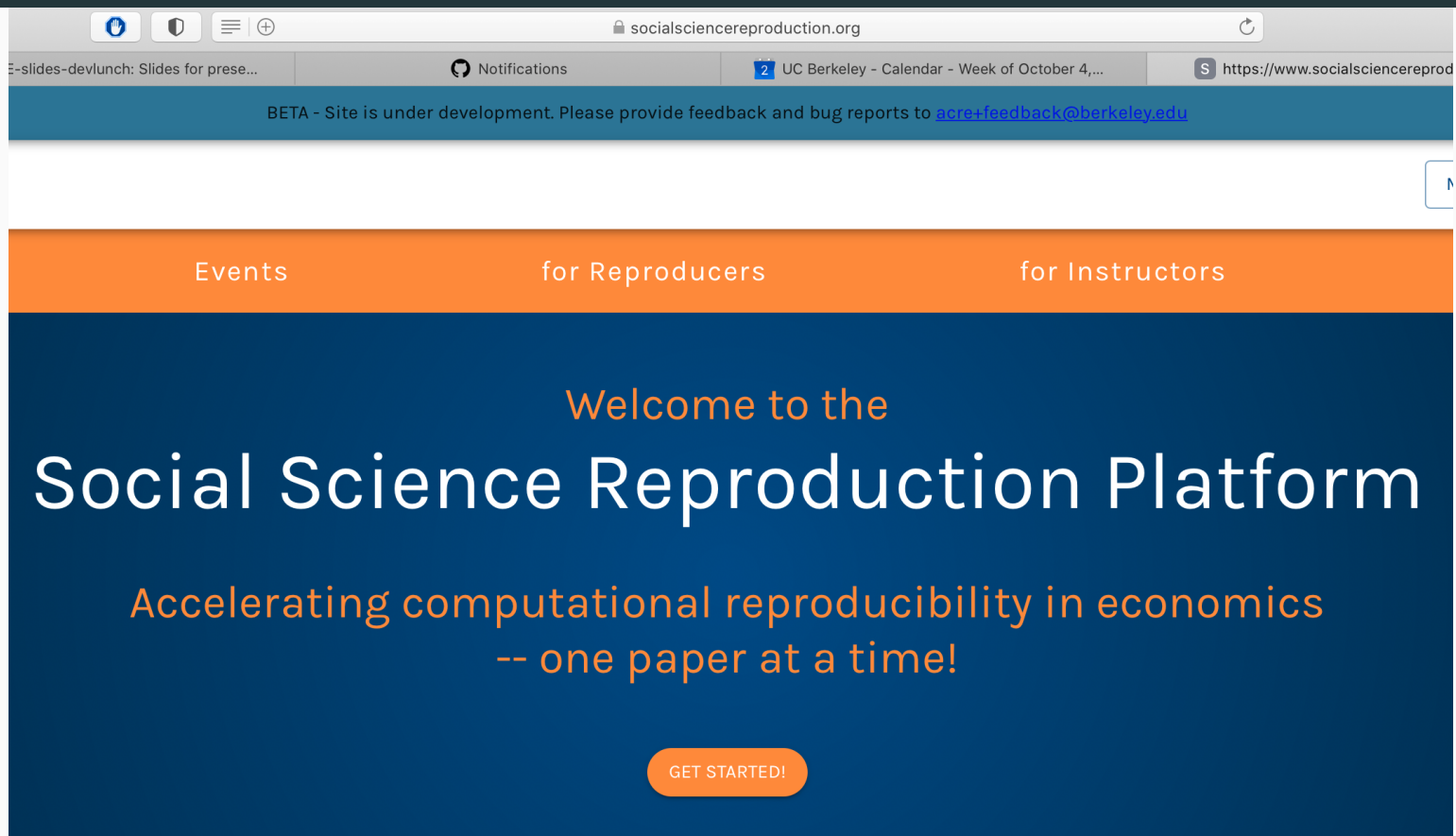This browser does not support PDFs. Please download the PDF to view it: Download PDF.

Easy to grade: report 1

This browser does not support PDFs. Please download the PDF to view it: Download PDF.

This browser does not support PDFs. Please download the PDF to view it: Download PDF.

Table of Contents

socialsciencereproduction.org

socialsciencereproduction.org

BETA - Site is under development. Please provide feedback and bug reports to acre+feedback@berkeley.edu

Events   for Reproducers   for Instructors

Welcome to the

Social Science Reproduction Platform

Accelerating computational reproducibility in economics
-- one paper at a time!

GET STARTED!

Purpose

On the Social Science Reproduction Platform, you can record and review **verifications and improvements** to the **computational reproducibility** of published social science work.

This open source platform was developed by the Berkeley Initiative for Transparency in the Social Sciences (BITSS) in collaboration with the American Economic Association Data Editor.

Acknowledgements

Arnold Ventures

Everybody who has participated in the pilots so far:

Ted Miguel's Graduate Development Economic Course (2019, 2020) - UC Berkeley

Dina Pomeranz undergraduate thesis for Marc Richter - University of Zurich

Slides template: Grant McDermott.