Replication Package for: Solar Eclipses and the Origins of Critical Thinking and Complexity.

Date 2023-03-06

Overview

This replication package accompanies Litina and Roca-Fernández. (forthcoming). "Solar Eclipses and the Origins of Critical Thinking and Complexity". The Economic Journal. DOI.

This articles uses multiple data sources that are described below. 20 interconnected Stata files run the code to generate the results for the 20 figures and 19 tables in the paper, including those in the Appendices. All the required data are available in open and Stata formats. The replicator should expect the code to run for about 10 minutes.

Computational requirements

The replicator is expected to have an up-to-date version of Stata 17 and Python (version 3.10 or more recent). The code was thoroughly tested on Linux with kernel version 6.3.7. Stata do-files should run on other operating systems, except for calls to the program `echo` that is only natively available on Linux, MacOS, and BSD systems. Lacking it does not hinder the replication exercise. Note that running the Python code on Windows may require some changes to the scripts. There are no special requirements in terms of hardware.

Authors

- Anastasia Litina
- Èric Roca Fernández

Data availability and provenance statements

Statement about rights

The author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of availability

- All data is publicly available, although accessing some may require a registration.
    - Downloading the dataset by Ashraf and Galor (2011) requires registering at OpenICPSR (free of charge, immediate registration).

Details on each data source

- The paper uses the Ethnographic Atlas (Murdock, 1967). We used Pat Gray's version, which includes additional variables related to climate. Unfortunately, the website hosting it is no longer accessible: see webarchive. More in detail, we downloaded the file linked under "Rdata format version of Ethnographic Atlas". We provide a copy of this file.
- The paper uses the Standard Cross-Sectional Sample (Murdock and White, 1969). We use the version provided by the d-place project, accessible here. The data is distributed under the Creative Commons Attribution 4.0 International License. In particular, this script harmonizes ethnic group names.
- The paper uses data from Eff and Abhradeep (2013).
- The paper uses data from Ashraf and Galor (2011). We use the data from the authors' replication files. The data is distributed under the Creative Commons Attribution 4.0 International Public License. For convenience, we renamed the file `20081371_Dataset.dta` as `Ashraf_Galor__2011.dta` but, otherwise, we use it as provided. Any changes made to the data are documented in the corresponding `4-Ashraf_Galor_Preparation.do`.
- The paper uses data from Michalopoulos and Xue (2021) on folklore topics. This can be downloaded from here. These data is distributed under the Creative Commons 1.0 Universal Public Domain. We use two files:
  - `motifs_ea_groups.dta` to build our variable about how eclipses are explained in the folklore
  - `concept_frequencies_ea_groups.dta` from which we retrieve the relevance of several concepts in the folklore, as well as aggregate information (number of tales and bibliography consulted).
- The paper uses the Seshat databank as derived in Miranda and Freeman (2020). These data is available here. The data is distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License.
- The paper uses data from the Wikidata project, distributed under the Creative Commons 1.0 Universal Public Domain License. Because these data changes over time, we provide the database that we retrieved at the time of writing.
- The paper uses georeferenced information about eclipses visibility from Jubier for solar eclipses –anular and total– and lunar eclipses. The data is freely available and can be used in non-commercial applications.
- We use several datasets from Fenske (2014). First, ethnic homeland characteristics (elevation, ruggedness, malaria, latitude and longitude) can be obtained here The data is freely available and requires a proper citation. However, the underlying maps that we use were not made public and we obtained these through personal communication. The files `calories_centroids.csv` and `geography.csv` are derived from these maps (`allmerge_2011_01.shp`). Geography is just a dump of the database embeded in in the shapefile. Calories are obtained by

averaging potential caloric yields before 1500 from Ashraf and Galor (2011) in qGIS using the `zonal statistics` plugin and computing its centroids, respectively; ethnicities are renamed using the information in `merges_by_Anastasia`.

- We employ other data sets throughout the paper:
    - Information on cloud coverage from Wilson and Jetz (2016). We used the Mean Annual tiff file. This data is distrubuted under the Creative Commons Attribution-NonCommercial 4.0 International License.
    - Information on lightning from the LIS/OTD 0.5 Degree High Resolution Monthly Climatology (HRMC) from NSSTC Thes data is produced by a USA Government agency and, thus, under the Public Domain.
    - Information on volcanoes from GVP (2013), available here and here Use of the former falls under Fair Use and the latter is distributed under the GNU General Public License v3.0.
    - Data on coastlines and rivers from Natural Earth - Coastline v4.1.0 and Natural Eath - Rivers v5.0.0 These data are Public Domain from Kelso and Patterson (2012).
    - Country boundaries are downloaded from the GADM project
    - Information on fault lines from Ahlenius H. available here. These data is available under the Open Data Commons Attribution License.
    - Information on potential caloric yields from Galor and Özak available here The data is distributed under the Creative Commons Attribution-ShareAlike 4.0 International License. We use two raster files: pre1500AverageCaloriesNo0.tif and post1500AverageCaloriesNo0.tif
    - Information on terrain altitude and ruggedness is derived from GMTED 2010 database Thes data is produced by a USA Government agency and, thus, under the Public Domain. We use the file Mean Statistic 30-arc seconds to compute average altitude and the file Standard Dev. Statistic 30-arc seconds to compute ruggedness. More in detail, for a given area, we measure ruggedness by averaging the different values of the standard deviation in altitude.
    - Distance to Addis Abeba is computed by measuring the travelling distance from ethnic homeland centroids to the coordinates of Addis Abeba. We used Özak (2010) Human Mobility Index HMI raster to derive optimal travel paths. The data is distributed under the Creative Commons Attribution-ShareAlike 4.0 International License. We forced all American ethnic groups to travel through the Bering strait. Computations were made using the `r.drain` GRASS package in qGIS version 2.14. For this reason, we cannot provide the replication code for this exercise but we explain below the necessary steps.
        * First, we reprojected the original HMI to EPSG:3832 to ensure a connection between the Americas and Asia through teh Bering strait. `gdalwarp -t_srs EPSG:3832 -r near -multi`

```
-of GTiff -co COMPRESS=DEFLATE -co PREDICTOR=2 -co
ZLEVEL=9 HMI.tif HMI_Reprojected.tif
```
* We create a temporary layer containing a point that represents Addis Ababa at coordinates 9.027, 38.736
* We use the GRASS plugin `r.cost` to compute the cumulative cost from each cell to Addis Ababa, where we assign a large cost to null cells, forcing travel over land (this step is requires plenty of memory, we cut the world in different regions and processed them separately).
* We used the GRASS plugin `r.drain` and the direction map created before to find the optimal route.

List of tables and figures

The provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

Description of programs and code

Instructions to generate the data and replicate the results

- The provided codes expect some folders to exist. Please ensure that, within the root folder (denoted `./` here) from which codes will be executed, the following folder structure exist:[1]
  - `./Latex/Figures/Robustness`
  - `./Latex/Tables`
- The code consists of two distinct parts:
  - The creating of several datasets, including their process, is delegated to a series of Python scripts that we detail below. For the purpose of replication, running these files is **not** required because we provide their outputs for convenience. As a matter of fact, some datasets such as Wikidata, are subject to continuous change.
  - Nevertheless, if the replicator wishes to fully recreate all the databases they can contact the authors. Therefore, to replicate the results of the paper the replicator should rely on the provided files. Four Python scripts generate Figures that appear in the paper.
    * We run the different scripts using Python 3.10.9. Required packages include pandas, geopandas, rasterio, shapely, scipy, fiona, statsmodels, seaborn as well as their dependencies. These can be installed from the distribution's repository or Python PyPy repository.
  - The data analysis and the creation of the tables and most of the figures is done in Stata with the help of the do-files provided in this

4

replication package.

  * To run them, *first adjust the global variable `base_path` in line 6 to point to the root of the downloaded replication folder.*
  * The file `1-Preamble.do` is enough to replicate the entire paper, including generating the Tables and most Figures (see above for Figures generated in Python). Tables are exported in LaTeX format.
  * In this paper, we used Stata 17, running under Linux 6.3.7.
  * The numbers referred in the text are computed within Stata and exported to a csv-like file named `exported_values.txt` inside the ./Latex folder. However, we export these using a shell command (`!echo`) that is only available in Linux.
  * We use external programs: estout, acreg (ranktest, hdfe), reghdfe (ftools), coefplot, binscatter, binsreg, xlincom, and speccurve (line 24 in `1-Preamble.do` provides precise instructions on how to install it); together with their dependencies. The do-files include the code to install the required packages.

Code organization

- The different programs are organized in folders:
  - `Scripts` contain the Python scripts we use to process georeferenced sources and plot some results. We used `Spyder` to run these scripts. Because of differences between packages, notably `matplotlib`, we recommend exploring the plots within `Spyder`.[2]
    * `12-Create_maps.py` generates Figure 1
    * `13-Plot_discoveries.py` generates Figure C.6.
    * `15-Plot-robustness.py` generates Figures C.10. to C.14. (requires running the Stata files first).
    * `14-Plot_regressions.py` generates Figures 2, C.8. and C.9. (requires running the Stata files first).
  - `Stata` contains the Stata files that generate all the remaining Figures and all the Tables.

Running order for the replication of the results and figures in the paper

1. Move inside the `Stata` directory
2. Run the file `1-Preable.do`, adapting the paths as needed.
3. Move to the `Scripts` folder: the following scripts use databases created in Stata, so it is imperative to run the Stata files before.
4. Create the maps and some additional figures by running, adapting the paths as needed: **Note for Windows users running the Python scripts**: when modifying the variable `base_path`, be sure to use forward slashes "/" instead of backwards slashes "\". For instance, instead of writing `C:\Users\` you should write `C:/Users/`.

1. `12-Create_maps.py`
2. `13-Plot_discoveries.py`
3. `14-Plot_regressions.py`
4. `15-Plot-robustness.py`

List of tables and figures

The provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

Filenames and Tables

Table	Filename
Table 1	Table_Development.tex
Table 2	Table_Proxies_Development.tex
Table 3	Table_Proxies_Human_Capital.tex
Table 4	Table_Proxies_Technology.tex
Table 5	Table_Proxies_Curiosity.tex
Table 6	Table_Religion.tex
Table 7	Table_Placebo.tex
Table C.1	Table_Development_15-Plot-robustness.tex
Table C.2	Table_Proxies_Development_15-Plot-robustness.tex
Table C.3	Table_Proxies_Human_Capital_15-Plot-robustness.tex
Table C.4	Table_Proxies_Technology_15-Plot-robustness.tex
Table C.5	Table_Proxies_Curiosity_15-Plot-robustness.tex
Table C.6	Table_Area.tex
Table C.7	Table_Area_Folklore.tex
Table C.8	Table_Quantile.tex
Table C.9	Table_Int_1.tex
Table C.10	Table_Int_2.tex
Table C.11	Table_Int_3.tex
Table C.12	Table_Folklore_Fraction.tex
Table C.13	Summary_Statistics.tex
Figure 1	Total_Eclipses.png
Figure 2	V31.pdf and V30.pdf
Figure 3	Eclipses_Science.pdf
Figure 4	Eclipses_Religion.pdf
Figure C.1	Files named Random_Eclipses_*.pdf[3]
Figure C.2	group_1.pdf
Figure C.3	group_2.pdf
Figure C.4	group_3.pdf
Figure C.5	EA_Over_time.pdf and Folklore_Over_time.pdf
Figure C.6	Discoveries_Greece.pdf and Discoveries_India.pdf

Table	Filename
Figure C.7	eclipses_distribution.png
Figure C.8	V33.pdf, V66.pdf, V90.pdf, Games.pdf, Writing.pdf and Explanation.pdf
Figure C.9	Calendar.pdf, Tasks.pdf, Technology.pdf, Thinking.pdf, Curious.pdf and Eclipse.pdf
Figure C.10	Files named 10_*.pdf
Figure C.11	Files named 25_*.pdf
Figure C.12	Files named 50_*.pdf
Figure C.13	Files named 100_*.pdf
Figure C.14	Files named 150_*.pdf
Figure C.15	Files named binsreg_bivariate_*.pdf and binsreg_multivariate_*.pdf
Figure C.16	Science_binscatter_1.pdf, Science_binscatter_2.pdf, Science_binscatter_Limited_1.pdf and Science_binscatter_Limited_2.pdf

Code references

- Table 1:
  - `Stata/2-EA.do` lines 7–81
  - `Stata/3-Seshat.do` lines 74–105 and 192–214
  - `Stata/6-Tables.do` lines 1–43
- Table 2:
  - `Stata/2-EA.do` lines 110–224
  - `Stata/3-Seshat.do` lines 74–105 and 192–214
  - `Stata/6-Tables.do` 45–89
- Table 3:
  - `Stata/2-EA.do` lines 110–224 and lines 231–275
  - `Stata/3-Seshat.do` lines 74–105
  - `Stata/6-Tables.do` lines 91–138
- Table 4:
  - `Stata/2-EA.do` lines 110-224
  - `Stata/3-Seshat.do` lines 74–105
  - `Stata/4-Ashraf_Galor.do`
  - `Stata/6-Tables.do` lines 140–183
- Table 5:
  - `Stata/2-EA.do` lines 231–275
  - `Stata/5-Wikipedia.do` lines 7–25
  - `Stata/6-Tables.do` lines 185–228
- Table 6:
  - `Stata/2-EA.do` lines 278–283 and lines 231–275
  - `Stata/5-Wikipedia.do` lines 7–25
  - `Stata/6-Tables.do` 230–272
- Table 7:

- – `Stata/2-EA.do` lines 231–275
- – `Stata/6-Tables.do` lines 274–308
- Table C.1:
  - – `Stata/2-EA_robustness.do` lines 2–24
  - – `Stata/3-Seshat_robustness.do` lines 80–111
  - – `Stata/6-Tables_robustness.do` lines 1–56
- Table C.2:
  - – `Stata/2-EA_robustness.do` lines 2–24
  - – `Stata/3-Seshat_robustness.do` lines 80-111
  - – `Stata/6-Tables_robustness.do` lines 58–115
- Table C.3:
  - – `Stata/2-EA_robustness.do` lines 2–24 and lines 29–45
  - – `Stata/3-Seshat_robustness.do` lines 80-111
  - – `Stata/6-Tables_robustness.do` lines 117–174
- Table C.4:
  - – `Stata/2-EA_robustness.do` lines 2–24
  - – `Stata/3-Seshat_robustness.do` lines 80-111
  - – `Stata/4-Ashraf_Galor_robustness.do`
  - – `Stata/6-Tables_robustness.do` lines 176–233
- Table C.5:
  - – `Stata/2-EA_robustness.do` lines 29–45
  - – `Stata/5-Wikidata_robustness.do`
  - – `Stata/6-Tables_robustness.do` lines 235–294
- Table C.6:
  - – `Stata/2-EA_Area.do` lines 8–95
  - – `Stata/6-Tables_Area.do` lines 1–70
- Table C.7:
  - – `Stata/2-EA_Area.do` lines 8–95
  - – `Stata/6-Tables_Area.do` lines 72–142
- Table C.8:
  - – `Stata/2-EA_robustness.do` lines 49–63
  - – `Stata/6-Tables_Area.do` lines 335–368
- Table C.9:
  - – `Stata/2-EA_Area.do` lines 106–130
  - – `Stata/6-Tables_Area.do` lines 145–209
- Table C.10:
  - – `Stata/2-EA_Area.do` lines 106–130
  - – `Stata/6-Tables_Area.do` lines 211–286
- Table C.11:
  - – `Stata/3-Seshat_robustness.do` lines 122–180
  - – `Stata/6-Tables_Area.do` lines 288–341
- Table C.12:
  - – `Stata/2-EA_robustness.do` lines 29–45
  - – `Stata/6-Tables_robustness.do` lines 296–329
- Figure 1:
  - – `Scripts/12-Create_maps.py`

- Figure 2:
  - `Scripts/14-Plot_regressions.py`
- Figure 3:
  - `Stata/5-Wikipedia.do` lines 34–67
- Figure 4:
  - `Stata/5-Wikipedia.do` lines 70–103
- Figure C.1:
  - `Stata/2-EA_robustness.do` lines 177–235
- Figure C.2:
  - `Stata/2-EA_Speccurve.do`
- Figure C.3:
  - `Stata/2-EA_Speccurve.do`
- Figure C.4:
  - `Stata/2-EA_Speccurve.do`
- Figure C.5:
  - `Stata/2-EA_Over_time.do`
- Figure C.6:
  - `Scripts/13-Plot_discoveries.py`
- Figure C.7:
  - `Scripts/7-Compute_eclipse_incidence.py`
- Figure C.8:
  - `Scripts/14-Plot_regressions.py`
- Figure C.9:
  - `Scripts/14-Plot_regressions.py`
- Figure C.10–C.14:
  - `Scripts/15-Plot-robustness.py`
- Figure C.15:
  - `Stata/2-EA-robustness.do` lines 151–165
- Figure C.16:
  - `Stata/5-Wikipedia.do` lines 109–127

References

- Ahlenius, H. (2014). World tectonic plates and boundaries. Github repository
- Ashraf, Q. and Galor, O. (2011). Dynamics and stagnation in the Malthusian epoch. American Economic Review, 101(5):2003–2041. Data
- Danielson, J.J., and Gesch, D.B. (2011), Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26 p. Data
- Eff, E. and Maiti, A. (2013). A measure of technological level for the Standard Cross-Cultural Sample. Working Papers 201302, Middle Tennessee State University, Department of Economics and Finance. The data appears in the paper.
- Fenske, J. (2014). Ecology, trade, and states in pre-colonial Africa. Jour-

nal of the European Economic Association, 12(3):612–640. Data

- Galor O. and Özak, Ö. (2016). The Agricultural Origins of Time Preference, American Economic Review, 106(10): 3064–3103. Data
- GADM (2019). GADM maps and data. Data
- Henderson, C. (2018). Demonstration of Jupyter notebook deployed with Binder. Github repository
- The v2.2 gridded satellite lightning data were produced by the NASA LIS/OTD Science Team and are available from the Global Hydrology Resource Center
- Jubier, X. M. (2019). Solar eclipses Accessed: 2019
- Jubier, X. M. (2019). Lunar eclipses. Accessed: 2019
- Kathryn R. Kirby, Russell D. Gray, Simon J. Greenhill, Fiona M. Jordan, Stephanie Gomes-Ng, Hans-Jörg Bibiko, Damián E. Blasi, Carlos A. Botero, Claire Bowern, Carol R. Ember, Dan Leehr, Bobbi S. Low, Joe Mc-Carter, William Divale, & Michael C. Gavin. (2021). D-PLACE/dplace-data: D-PLACE – the Database of Places, Language, Culture and Environment (v2.2.1) [Data set]. Zenodo.
- Kelso N. V. and Patterson T. (2012), Natural Earth. [Data set]. Natural Earth
- Michalopoulos, S. and Xue, M. M. (2021). Folklore. The Quarterly Journal of Economics, 136(4):1993–2046. Data
- Miranda L. and Freeman J. (2020). The two types of society: Computationally revealing recurrent social formations and their evolutionary trajectories. PLoS ONE, 15(5):e0232609. [Data set]: Seshat Accessed: 2019
- Murdock, G. P. (1967). Ethnographic atlas: A summary. Ethnology, 6(2):109–236. Downloaded from Archive
- Murdock, G. P. and White, D. R. (1969). Standard cross-cultural sample. Ethnology, 8(4):329–369.
- Global Volcanism Program, 2022. [Database] Volcanoes of the World (v. 4.6.7; 20 Mar 2018). Distributed by Smithsonian Institution, compiled by Venzke, E. 10.5479/si.GVP.VOTW5-2023.5.1
- Özak, Ömer (2010). The voyage of homo-economicus: Some economic measures of distance. Department of Economics, Brown University. Data
- Venzke, E., editor (2013). Global Volcanism Program. Smithsonian Institution.
- Wilson, A. M. and Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. PLOS Biology, 14(3):e1002415. Data
- Wikidata - A Collaborative Knowledge Base. Wikidata

---

1. Please ensure that the imputed path ends at the folder "3 replication package", and that the full path to this folder is copied. For instance, assuming the replicator uses Windows, the global variable `base_path` in line 6 of `1-Preamble.do` could look like `C:\Users\YourUsername\Downloads\Litina_Roca-Fernandez__2019/3`

replication package

2. This may arise from differences in the rendering backend between Linux and Windows.

3. Variable v30 is the variable tracking "Settlement Patterns", likewise V31 corresponds to "Population density", v33 to "Jurisdictional Hierarchy", v66 to "Class Stratification" and v90 to "Political Integration".