

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Connolly, Marie; Djamaldiev, Oumar; Munteanu, Andrei

Working Paper Replication of "Solar Eclipses and the Origins of Critical Thinking and Complexity"

I4R Discussion Paper Series, No. 153

Provided in Cooperation with: The Institute for Replication (I4R)

Suggested Citation: Connolly, Marie; Djamaldiev, Oumar; Munteanu, Andrei (2024) : Replication of "Solar Eclipses and the Origins of Critical Thinking and Complexity", I4R Discussion Paper Series, No. 153, Institute for Replication (I4R), s.l.

This Version is available at: https://hdl.handle.net/10419/302897

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

INSTITUTE for

No. 153 I4R DISCUSSION PAPER SERIES

Replication of "Solar Eclipses and the Origins of Critical Thinking and Complexity"

Marie Connolly Oumar Djamaldiev Andrei Munteanu

September 2024



I4R DISCUSSION PAPER SERIES

I4R DP No. 153

Replication of "Solar Eclipses and the Origins of Critical Thinking and Complexity"

Marie Connolly¹, Oumar Djamaldiev¹, Andrei Munteanu¹

¹Université du Québec à Montréal/Canada

SEPTEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and metascientific work in the social sciences. Provided in cooperation with EconStor, a service of the <u>ZBW – Leibniz Information Centre for Economics</u>, and <u>RWI – Leibniz Institute for Economic Research</u>, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur University of Ottawa Anna Dreber Stockholm School of Economics Jörg Ankel-Peters *RWI – Leibniz Institute for Economic Research*

E-Mail: joerg.peters@rwi-essen.de RWI – Leibniz Institute for Economic Research Hohenzollernstraße 1-3 45128 Essen/Germany www.i4replication.org

Replication of "SOLAR ECLIPSES AND THE ORIGINS OF CRITICAL THINKING AND COMPLEXITY"*

Marie Connolly[†], Oumar Djamaldiev[‡], Andrei Munteanu[§]

May 2024

Abstract

Overall, the results of this study are replicable using the clean dataset provided by the authors. However, the process of transforming the raw data into this clean dataset is at times unclear, and certain essential codes are missing from the replication materials, which may hinder complete replication. Despite these challenges, the paper's findings remain robust to a small number of robustness checks that we conducted, providing confidence in the reliability and stability of the results. Moving forward, efforts to enhance transparency in data transformation steps and inclusion of all necessary codes would facilitate more seamless replication and validation of the study's outcomes.

^{*}Litina, A., & Roca Fernández, È. (2024). Solar Eclipses and the Origins of Critical Thinking and Complexity. *The Economic Journal*, 134(660), 1545-1577. https://doi.org/10.1093/ej/uead117.

Data and code replication package: Litina, A., & Roca Fernández, È. (2023). Replication package for "Solar Eclipses and the Origins of Critical Thinking and Complexity". [Data set]. Zenodo. https://doi.org/10. 5281/zenodo.10408724

[†]Département des sciences économiques, ESG UQAM (Montréal, Canada) and CIRANO (Montréal). Email: connolly.marie@uqam.ca

[‡]Département des sciences économiques, ESG UQAM (Montréal, Canada). Email: djamaldiev.oumar@courrier.uqam.ca

 $^{^{\}S}$ Département des sciences économiques, ESG UQAM (Montréal, Canada). Email: munteanu.andrei@uqam.ca

1 Introduction

We reproduced and replicated the findings of Litina and Roca Fernández (2024). This report presents our observations based on this exercise. In summary, the replication package (Litina and Roca Fernández, 2023) effectively reproduced the tables and figures from the original paper. Nonetheless, transitioning from the raw data to the data included in the replication package presents challenges. A list of the data files included in the replication package clearly mentioning their sources and a proper reference would be very useful. It is important to acknowledge that certain data cannot be shared directly due to external constraints beyond the authors' influence. A dditionally, the absence of code for certain data transformations or scraping procedures utilized by the authors is another notable observation. This aspect could be addressed to enhance the transparency and reproducibility of the research findings.

2 Running the Replication Package

The organization of the replication package was commendable, featuring well-commented and structured code. Reproducing the results was straightforward; simply adjusting the path to the working directory and executing the main code led to successful computational reproduction. It's worth noting that, as specified by the authors, S tata 17 was necessary for running the package, indicating incompatibility with Stata 16.

However, the accompanying ReadMe file posed some challenges in terms of clarity and coherence. First, it was not always obvious to go from the verbose description of data sources in the Details on each data source section to the data files provided in the replication package. A simple table listing each data file, a long with its source, could do the trick. We also note that there is no clear separation between the input data sources (raw files) and the files generated throughout the analysis (intermediate or analysis files). Second, the ReadMe h inted at c ertain a nalysis steps that appeared to be missing or inadequately explained, warranting further elaboration to facilitate a complete understanding of the replication process.

3 From Raw to Final Data

In this section, we note some observations on a few selected data sources.

3.1 Ethnographic Atlas

The Ethnographic Atlas data were utilized for documenting various ethnic groups. The authors noted that the Ethnographic Atlas is referenced as an .Rdata file, which unfortunately is no longer accessible online and can only be retrieved through web archives. Regrettably, this specific file does not appear to be included in the replication package. Moreover, the authors did not supply the web archive link or detailed instructions on

how to retrieve this file. This absence of information poses a challenge in accessing and utilizing this critical dataset for replication purposes. Furthermore, we note that there appears to be some online sources for these data, but since they are not referenced directly, it is unclear whether they correspond to the data used in this paper (see for example https://d-place.org/contributions/EA).

3.2 Wikidata Data on Individual Scientists

This study relied on data sourced from Wikipedia, encompassing individuals born between 1600 BCE and 1800 CE, inclusive of their birth dates, birthplaces, and occupations. However, we encountered challenges in transitioning from the raw scraped data to the dataset utilized in the final analysis, which are outlined as follows:

- 1. The code employed for scraping Wikipedia using the Wikidata SPARQL endpoint was referenced only as a footnote in the main text of the paper, with no corresponding script included in the replication package.
- 2. The sample code provided in the footnote did not function correctly when applied directly to Wikidata, potentially due to formatting alterations that occurred during the process of copying and pasting from the paper.
- 3. The provided sample code specifically addressed individuals born in 1500, whereas the paper's footnote indicated the necessity to query data across the broad timeframe from 1600 BCE to 1800 CE. However, no guidance was offered on how to execute this range querying, particularly considering that SPARQL does not support traditional looping mechanisms.
- 4. Upon implementing our own iterative process in R to query Wikidata and retrieve a version of the dataset, we encountered discrepancies. Notably, certain variables related to occupations (specifically those tied to religion, science, etc.) were absent from the dataset provided by the authors, so we have no way to test how these variables were added to the dataset. Furthermore, it remains unclear how variables such as country and continent were derived from latitude and longitude coordinates within the raw data. Consequently, the replication package does not provide sufficient information to replicate the transformation process from raw data to the final dataset used in the study.

3.3 Eclipse Data

Understanding the exact construction process of the eclipse dataset used in the study presents challenges. Although the authors reference two sources (Jubier, 2019; Espenak and Meeus, 2006), it remains unclear how to access and download the specific data utilized in the paper. The appendix briefly mentions using a P ython script to scrape the path of totality and basic eclipse date information from maps sourced from these references; however, the script itself is not included in the provided materials. Moreover, the transformation from this raw eclipse data to a geo-database containing paths of totality lacks clarity. The authors state that they developed a geo-database that includes the path of totality and associated eclipse dates. This geo-database was then used to intersect eclipse paths with other geometric data such as ethnic homelands, polity territory, and individuals' birthplaces. Unfortunately, the specific code or methodology used to construct this geo-database is not provided in the replication package, which complicates efforts to replicate this aspect of the study's analysis.

4 Robustness

4.1 Eclipse Exposure at the Ethnic Level

We conducted several robustness checks by estimating alternative specifications similar to the initial columns in Table 4.1. One challenge we encountered was the presence of unnamed variables (e.g., "v31") in the dataset, which complicated intuitive specification changes.

A notable issue we identified with the clipse exposure variable is twofold: i) it exhibits numerous missing values, including significant ethnic groups like the Chinese and Romans, and ii) there exists a mechanical correlation between eclipse exposure and the size of countries (e.g., Russians and Brazilians showing high eclipse exposure due to the large surface area of their respective countries, whereas many Pacific Island ethnic groups or African ethnicities may not). We note however that the controls include indicators for the decile of land area, thus using variation in eclipse exposure within decile of land mass area.

Nonetheless, to mitigate potential outlier effects, we applied a rule of t humb by excluding cases with more than 700 eclipses over the past 4,000 years. The results can be seen in column (1) of Table 4.1 below. Despite this adjustment, the results remained statistically significant at a 5% confidence level.

A second robustness check we did was to use standardized number of eclipses instead of a logarithmic transformation. The results also demonstrating robustness to this adjustment (see column (2)).

Furthermore, we replaced the ordered probit analysis with an ordinary least squares (OLS) regression, and the results remained robust under this alternative model specification (positive c oefficient, statistically different from zero). Finally, we added the log of the total number of lunar eclipses as an additionnal regressor (column (4)), and again found the main results to be robust. A summary of these findings is presented in Table 4.1.

4.2 Eclipse Exposure at the Individual Level

The absence of code that links the scraped-data from Wikidata to the eclipse data for calculating eclipse exposure poses significant challenges for replicating and modifying

Table 1: Robustness Checks				
	(1)	(2)	(3)	(4)
	drop outliers	std	ols	add lunar
main				
$eclipses_log$.48		.86	.53
	(.24)		(.35)	(.29)
	[.04]		[.02]	[.07]
$std_{eclipses}$.29		
		(.09)		
		[.00]		
N	464	466	466	466

Source: Authors' computations based on data in Litina and Roca Fernández (2023). Note: This table presents the estimated coefficients on eclipses (in log, columns (1), (3) and (4), and standardized (column (2)), replicating results from column (2) of Table 1 in Litina and Roca Fernández (2024). Standard errors in parentheses, association p-values in brackets.

the additional variables such as occupational characteristics and eclipse exposure. This limitation makes it difficult to : i) Ve rify how these supplemental variables (e.g., occupational attributes, eclipse exposure) were specifically coded in the dataset and ii) Conduct placebo tests or implement modifications to assess alternative scenarios, such as replacing exposure to eclipses during ages 5 to 15 with exposure during ages 25 to 35, exposure levels of 0 to 5, or considering eclipses occurring before individuals' births.

Unfortunately, the missing code impedes our ability to carry out such analyses or modifications e ffectively. The inclusion of this critical linking code would greatly enhance the transparency and reproducibility of the study's methodology, facilitating further investigation and experimentation with the dataset.

Being unable to estimate models based on alternative eclipse exposure scenarios due to the missing data-creation steps in the replication package, we opted to conduct placebo tests by generating random eclipses within each country based on the number of eclipses recorded for that country in the dataset provided by the authors. This approach allowed us to assess the robustness of the models by seeing if a randomlygenerated eclipse exposure variable would yield null results (which is what one would expect from a randomly-generated X variable).

Specifically, w e s imulated 1,000 d ifferent ec lipse di stributions an d es timated the models corresponding to columns 3 and 4 of Table 5 in Litina and Roca Fernández (2024). Our findings are presented graphically in Figures 1 (corresponding to column (3) of Table 5) and 2 (column (4)). We generally find n ull r esults, w hich provide reassurance regarding the validity of the findings of Litina and Roca Fernández (2024). In the vast majority of cases, the estimated coefficient on random ec lipse exposure is not statistically different from zero.



Figure 1: Placebo Tests for Table 5: Column 3

Source: Authors' computations based on data in Litina and Roca Fernández (2023). Note: This figure presents the 1000 estimated coefficients on randomly-generated eclipse exposure, replicating results from column (3) of Table 5 in Litina and Roca Fernández (2024). The draws are sorted by the size of the estimated coefficient, and the figure shows the 95% confidence intervals.

5 Conclusion

In summary, the results of the study are replicable using the clean dataset provided by the authors. However, the process of transforming the raw data into this clean dataset is at times unclear, and certain essential codes are missing from the replication materials, which may hinder complete replication.

Despite these challenges, the paper's findings r emain r obust t o a s mall number of robustness checks that we conducted, providing confidence in t he r eliability and stability of Litina and Roca Fernández' results. Moving forward, efforts t o enhance transparency in data transformation steps and inclusion of all necessary codes would facilitate more seamless replication and validation of the study's outcomes.



Figure 2: Placebo Tests for Table 5: Column 4

Source: Authors' computations based on data in Litina and Roca Fernández (2023). Note: This figure presents the 1000 estimated coefficients on randomly-generated eclipse exposure, replicating results from column (4) of Table 5 in Litina and Roca Fernández (2024). The draws are sorted by the size of the estimated coefficient, and the figure shows the 95% confidence intervals.

References

- Espenak, F. and Meeus, J. (2006). Five millenium catalog of solar eclipses-1999 to+ 3000 http://eclipse.gsfc.nasa.gov/SEcat5/SEcatalog. NASA Goddard Space Flight Center.
- Jubier, X. M. (2019). Solar eclipses, http://xjubier.free.fr/en/site_pages/Solar_ Eclipses.html.
- Litina, A. and Roca Fernández, (2023). Replication package for "Solar Eclipses and the Origins of Critical Thinking and Complexity". https://doi.org/10.5281/zenodo. 10408724.
- Litina, A. and Roca Fernández, (2024). Solar Eclipses and the Origins of Critical Thinking and Complexity. *The Economic Journal*, 134(660):1545–1577.