

Nate Breznau

# Investigating the reproducibility of the social and behavioral sciences

+ a deep dive into implications for sociology



German Institute for  
Adult Education  
Leibniz Centre for  
Lifelong Learning

[www.die-bonn.de](http://www.die-bonn.de)

Cornell University  
Department of Sociology

May 1st, 2026  
3:30 – 4:45pm

Uris Hall G08



Deutsche  
Forschungsgemeinschaft



### Pre-Recorded Talk:

<https://youtu.be/kMwOwZaYYtE>



### Suggested Citation:

Breznau, Nate. 2026. “Investigating the reproducibility of the social and behavioral sciences + a deep dive into implications for sociology.” Invited talk at Cornell University Department of Sociology, May 1<sup>st</sup>, 3:30pm. Uris Hall. Zenodo Slide Sharing. <https://doi.org/10.5281/zenodo.19956336>

# Systemizing Confidence in Open Research and Evidence (SCORE)



Brian  
Nosek



# Systemizing Confidence in Open Research and Evidence (SCORE)

REPRODUCIBILITY

*Verifiability of numerical results*

ROBUSTNESS

*Alternative specifications*

REPLICABILITY

*New sample, same experiment*

# Systemizing Confidence in Open Research and Evidence (SCORE)

REPRODUCIBILITY

HUMAN v. MACHINE  
ASSESSMENTS

ROBUSTNESS

CREDIBILITY

REPLICABILITY

# The SCORE Project



Miske + 127 authors (2026). “Investigating the **reproducibility** of the social and behavioural sciences.” *Nature* 652:126–134. <https://doi.org/10.1038/s41586-026-10203-5>



Aczel + 490 authors (2026). “Investigating the analytic **robustness** of the social and behavioural sciences.” *Nature* 652:135–142. <https://doi.org/10.1038/s41586-025-09844-9>



Tyner + 291 authors (2026). “Investigating the **replicability** of the social and behavioural sciences.” *Nature* 652:143–150. <https://doi.org/10.1038/s41586-025-10078-y>



Abatayo + 85 authors (2026). “Assessments of credibility in the social and behavioural sciences.” Preprint on *MetaArXiv*. [https://osf.io/preprints/metaarxiv/7u58q\\_v1](https://osf.io/preprints/metaarxiv/7u58q_v1)



Rajtmajer + 39 authors (2026). “Innovations in machine assessment of replicability.” Preprint on *MetaArXiv*. [https://osf.io/preprints/metaarxiv/bdmte\\_v1](https://osf.io/preprints/metaarxiv/bdmte_v1)



Mody + 34 authors (2026). “Large-scale human predictions of the replicability of published social and behavioural science papers – a multi-study analysis.” Preprint on *MetaArXiv*. [https://osf.io/preprints/metaarxiv/vgyed\\_v1](https://osf.io/preprints/metaarxiv/vgyed_v1)

# Systemizing Confidence in Open Research and Evidence (SCORE)

*Today's Talk*

REPRODUCIBILITY

HUMAN ASSESSMENTS

ROBUSTNESS

CREDIBILITY

REPLICABILITY

## Study Design

---

- |   |                        |
|---|------------------------|
| 1. Sample journals in social & behavioral sciences:       | 62 journals            |
| 2. Identify papers with statistical tests:                | 3,000 papers (2009-18) |
| 3. Identify eligible sample of papers:                    | 600 papers             |
| 4. Get replication data (online / request):               | 146 papers (24.3%)     |
| 5. Identify main claims in papers for testing:            | 553 claims             |
| 6. Crowdsourced testing of computational reproducibility: | 114 analysts           |
| 7. R&R - investigate journal policies                     | Small volunteer team   |



# Process Reproducibility

---

## “Process reproducibility”

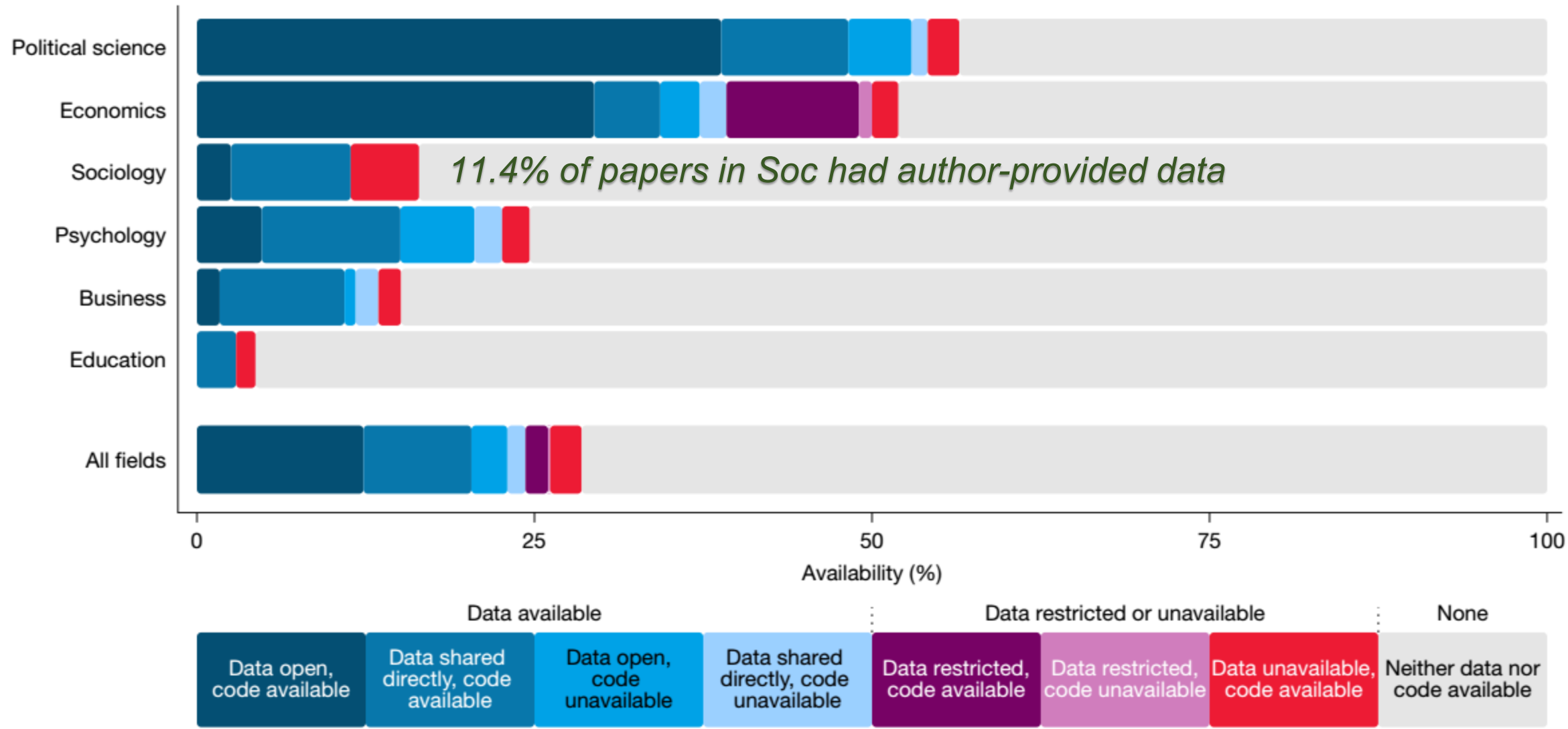
- Study is minimally reproducible, if data are available
- (Ideally code is also provided)

## “Author-provided data”

- Journal website
- Online repository
- Direct contact response

## “Reconstructed data” – from online sources

# Results – Data and Code Availability – Paper-Level



## Results – Measuring Computational Reproducibility

---

**Precisely reproduced** – all numerical outcomes identical to original

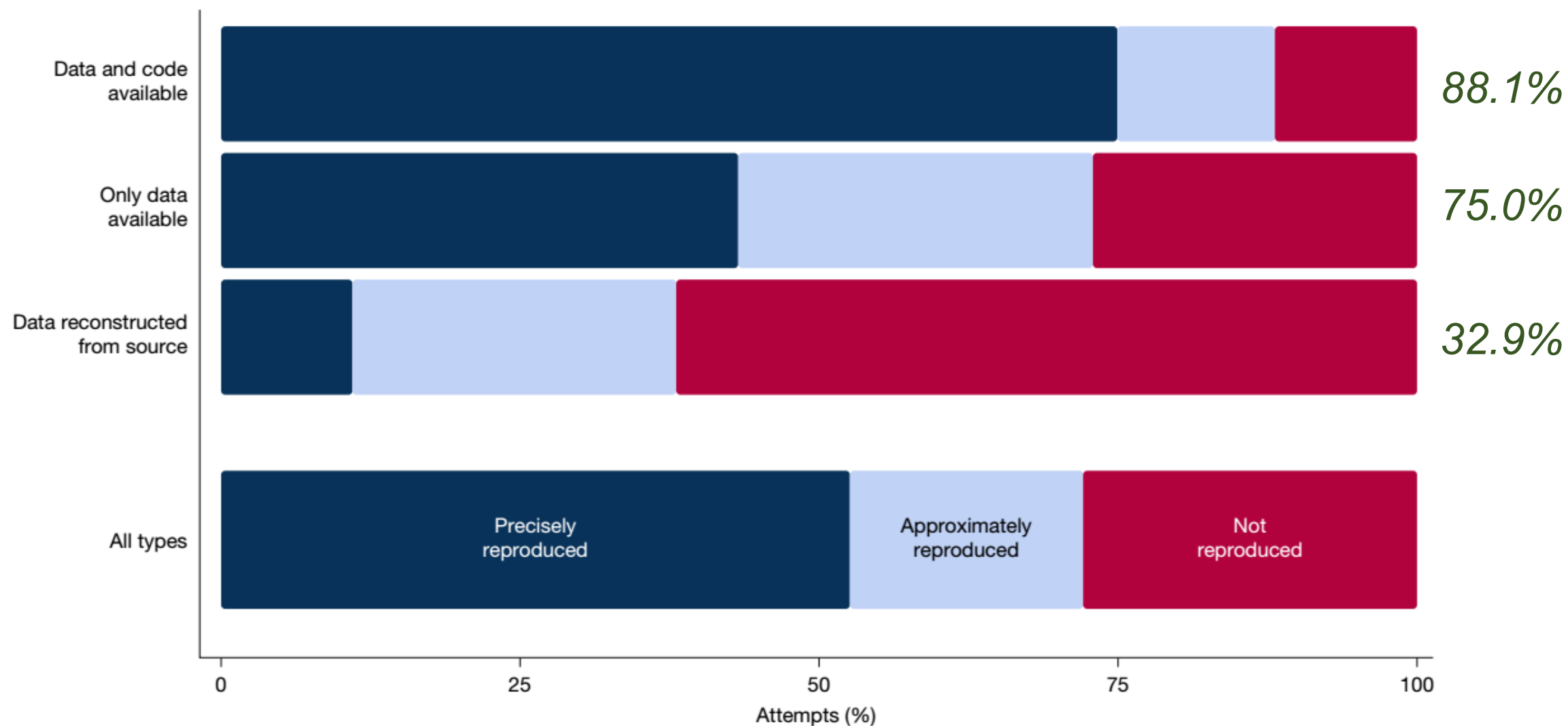
**Approximately reproduced** – at least one outcome not precise, but all outcomes within 15% of the original numerical finding

**Not reproduced** – neither of the above applies

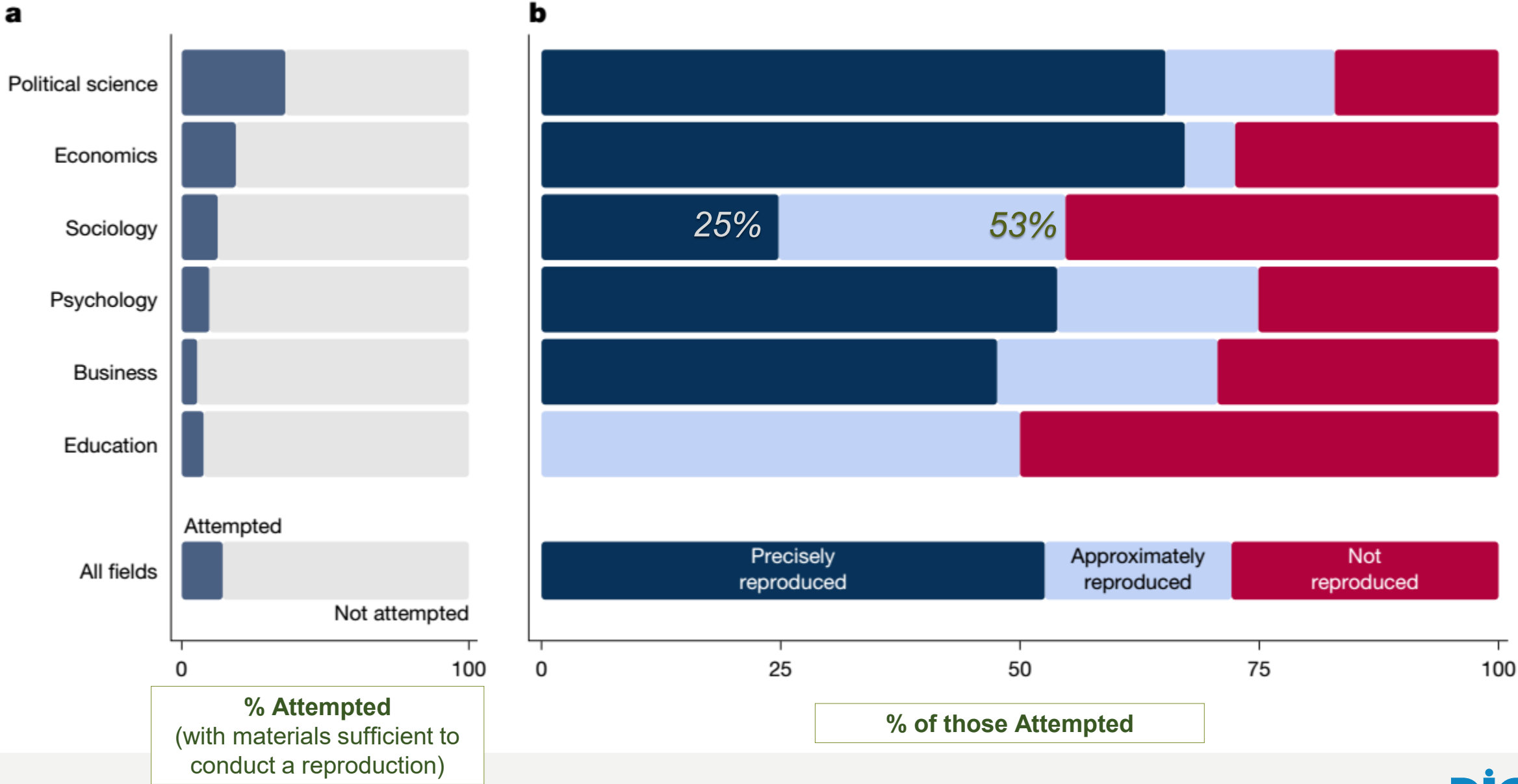
### **Achtung! This is complicated:**

- Each reproduction attempt occurs at the claim-level.
- Multiple claims per paper
  - inferential statistics require weighting
  - A paper with 5 claims, but only 1 with reproduction data, is weighted as 0.2 papers.
- Sociology had 79 eligible papers ( $\geq 1$  claim), but only 23 with materials sufficient for a reproduction ( $\geq 1$  claim), and only 21 had completed reproductions of at least 1 claim.
- Sociology had 53 claims in 21 papers, but because of the data availability, number of claims and weighting... = only 9 papers.

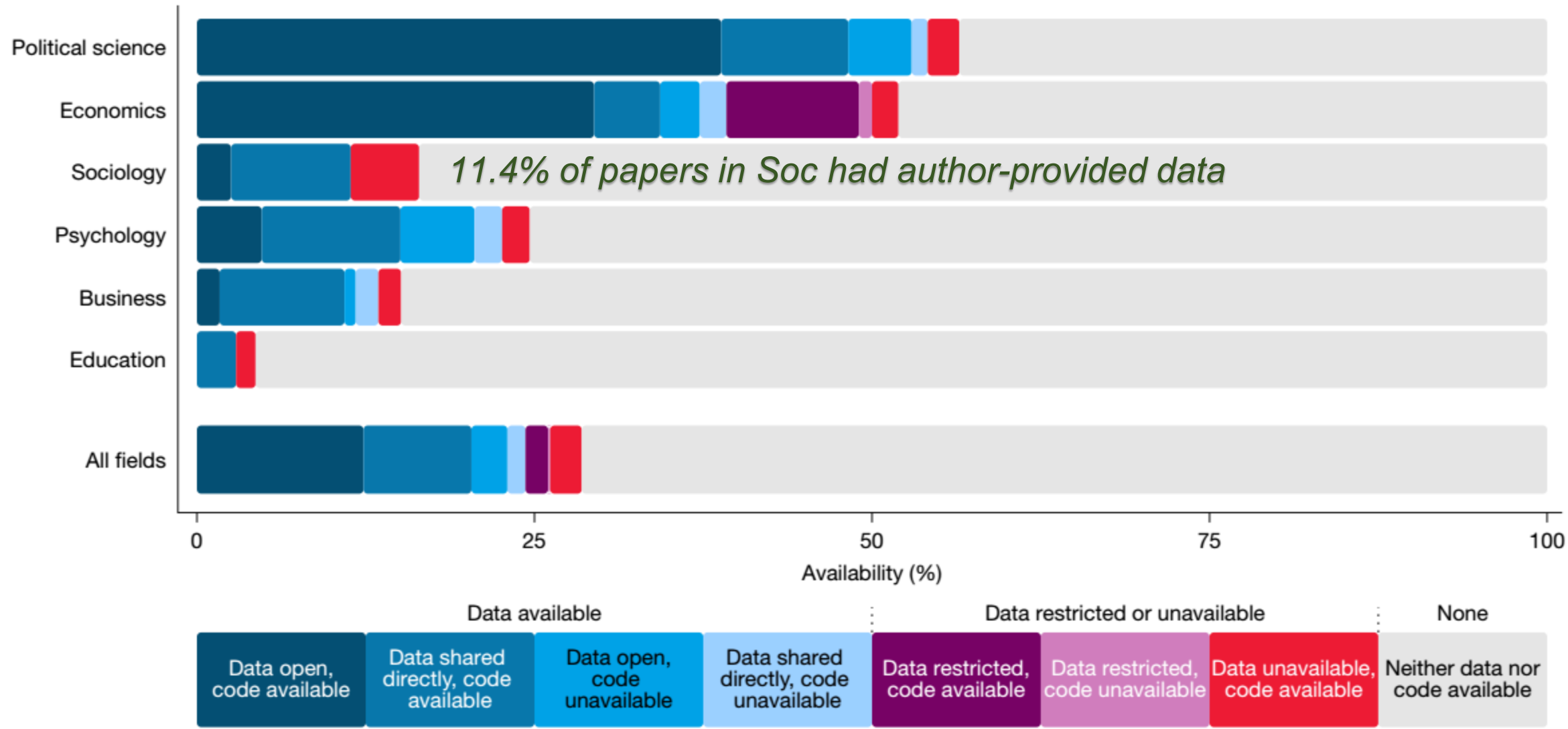
## Results – Computational Reproducibility – Claims-Level



# Results – Computational Reproducibility by Field – Claims-Level



# Results – Data and Code Availability – Paper-Level



## Conclusion from SCORE

---

### Reproducibility is precarious

- Only one-quarter of social and behavioral science studies from 2009-2018 shared their data
- Of these one-quarter, 51% were precisely reproducible. Sounds OK?, but...
- Only 17.5% of all studies were (just) approximately reproducible, when counting those without data as 'not reproducible'.



# Comparability of SCORE Results with Previous Reproducibility Research

---

Many studies investigating computational reproducibility

- Hardwicke et al. (2018): 35 studies in *Cognition* data available, 37% had at least one effect not statistically reproducible within 10%.
- Chang and Li (2022) & Gertler et al. (2018), data not always available - computational reproducibility rates between 14% and 43%.
- Krähmer, Schächtele, and Auspurg (2026). Review of studies using European Social Survey data: 35% of authors shared code. Of those, 51% numerically reproducible - overall exact reproducibility rate ~18%.
- Janz (2015): When American Journal of Political Science (AJPS) instituted mandatory reproducibility checks, the first 15 accepted papers all failed
- Eubank (2016), Liu & Salganik (2019), Trisovic et al. (2022) and Stockemer, Koehler, & Lentz (2018): Persistent data access and reproduction failures in social and political sciences

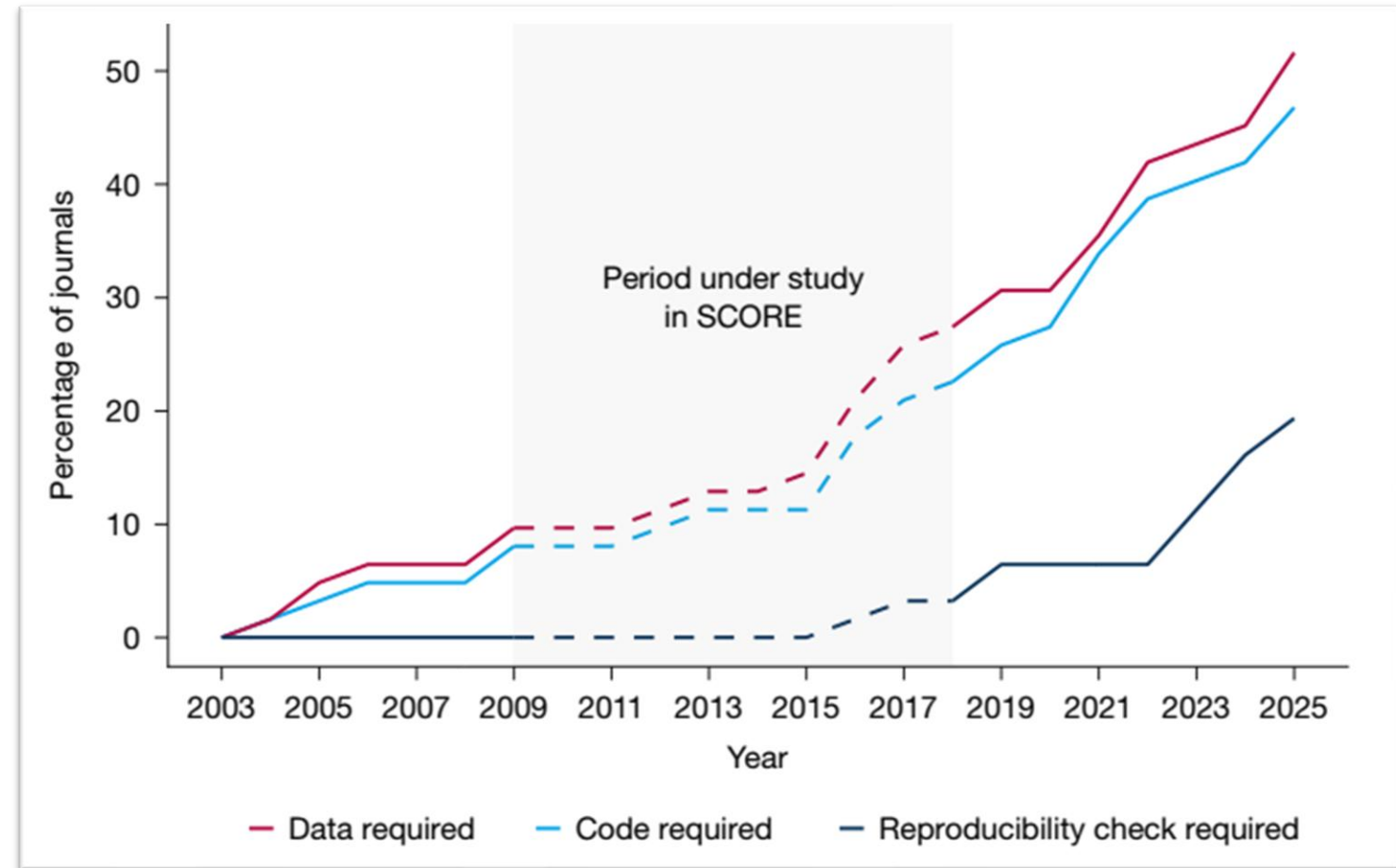
See review in: Breznau, Nate, Eike Mark Rinke, Alexander Wuttke et al. 2025. "The Reliability of Replications: A Study in Computational Reproductions." *Royal Society Open Science* 12(3):241038. doi:[10.1098/rsos.241038](https://doi.org/10.1098/rsos.241038).



# The Open Science Movement

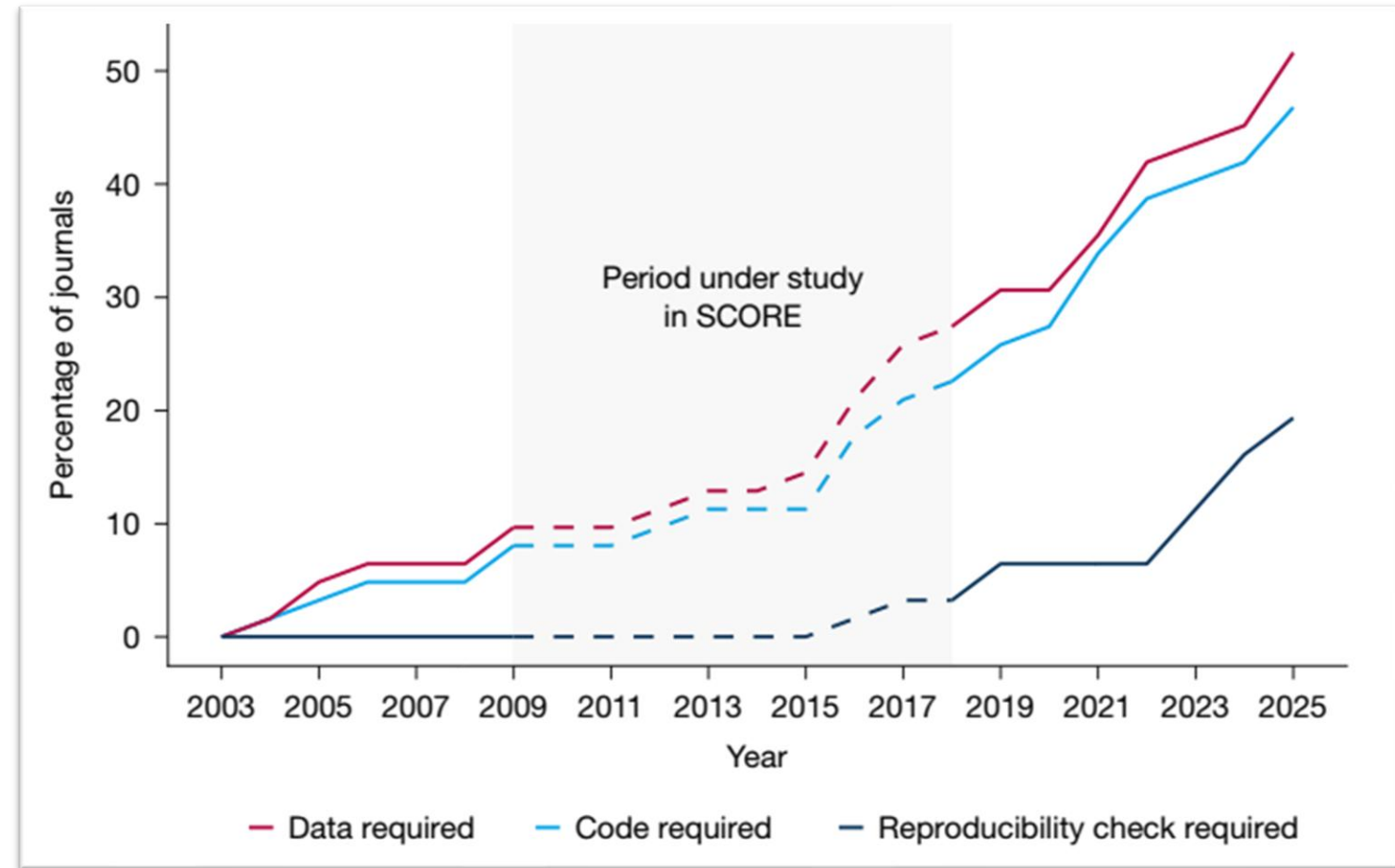
## Things are getting better quickly

- 88.1% of studies that share data and code were approximately reproducible
- Data and code sharing is becoming the norm



# Importance of Transparency Confirmed in Other Studies

- Breznau, Rinke and Wuttke (2025). 88 independent replications of the same study. Split design – those with code and materials 77% exact reproduction, those without code and materials 48%.
- Brodeur et al. (2026): Evaluation of 110 recently published articles in economics and political science journals with mandatory data and code sharing policies, 85%+ computationally reproducible.



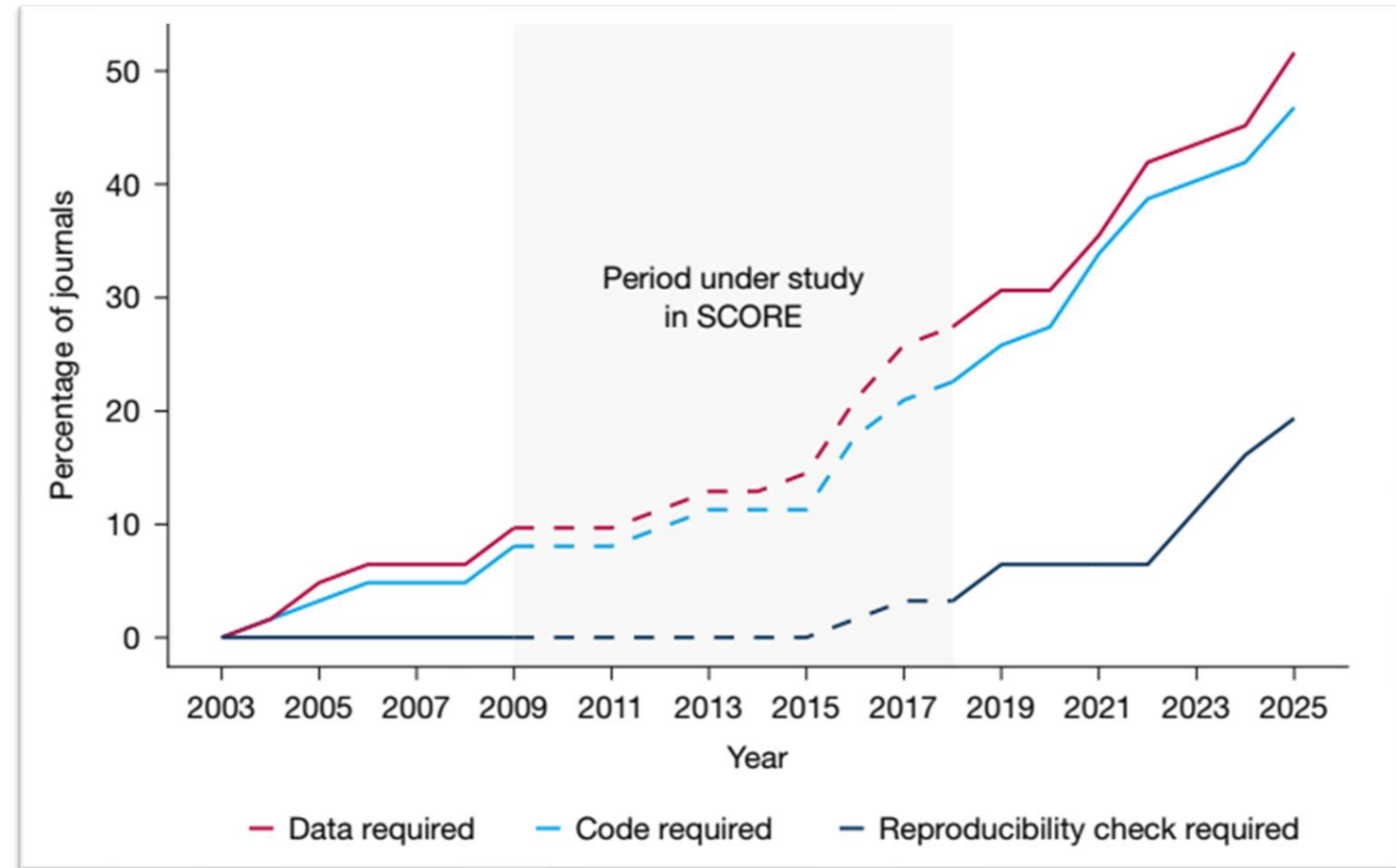
# The Open Science Movement

## BUT, sociology lags

*0% of sociology journals in the study require data sharing*

Poli Sci	77.8%
Econ	36.4%
Psych	26.7%
Business	8.3%
Education	0.0%
<b>Sociology</b>	<b>0.0%</b>

- *Law and Human Behavior*
- *Journal of Marriage and Family*
- *Criminology*
- *American Sociological Review*
- *American Journal of Sociology*
- *Demography*
- *Social Forces*
- *European Sociological Review*



# Replicability & Generalizability of SCORE

---

- |   |                        |
|---|------------------------|
| 1. Sample journals in social & behavioral sciences:       | 62 journals            |
| 2. Identify papers with statistical tests:                | 3,000 papers (2009-18) |
| 3. Identify eligible sample of papers:                    | 600 papers             |
| 4. Get replication data (online / request):               | 146 papers (24.3%)     |
| 5. Identify main claims in papers for testing:            | 553 claims             |
| 6. Crowdsourced testing of computational reproducibility: | 114 analysts           |
| 7. R&R - investigate journal policies                     | Small volunteer team   |

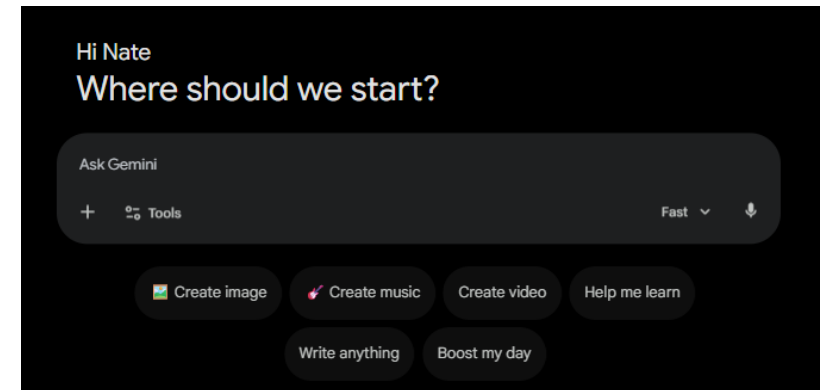
Sociology had 515 papers in the original sample,  
79 were investigated for data sharing.

***What about the other 436?  
& Have things changed over time?***

# Replicability & Generalizability of SCORE

---

1. Extract all sociology & political science studies
2. Prompt Gemini 3.1 Pro Web-Version, Paid Tier
3. Agentic mode – find articles, check for shared data, check for sharing statement
4. Human (Nate) somewhat random sample check for accuracy



## Sociologists' Data Sharing Behaviors at Parity

---

	Sociology	Political Science
<i>___ UNTESTED sample of papers from 2009-2018 (~2026)</i>		
Total papers	436	466
Data found	70 (16%)	68 (15%)
Data mention/found	169 ( <b>39%</b> )	162 ( <b>35%</b> )
<i>___ TESTED sample of papers from 2009-2018 (~2018)</i>		
Total papers	79	85
<b>ORIG: Data shared</b>	<b>23 (29%)</b>	<b>50 (59%)</b>
ORIG: Data online	1 (1%)	33 ( <b>39%</b> )
LLM: Data found	21 ( <b>27%</b> )	17 (20%)
LLM: Data mention/found	34 (43%)	33 (39%)

(~2026)

# Checking the LLM's Work & If Anything Changed Since 2018?

## SOCIOLOGY:

	LLM_Audit	
Original	FALSE	TRUE
FALSE	38	32
TRUE	7	2

Agreement: 50.63%

Title: The Scarring Effects Of Primary-Grade Retention? A Study Of Cumulative Advantage In The Educational Career  
DOI: 10.1093/sf/sou074. Orig: F | LLM: T – **SCORE replicator posted**

Title: Who Is The Residential Parent? Understanding Discrepancies In Unmarried Parents' Reports. DOI: 10.1111/jomf.12087.  
Orig: F | LLM: T – **posted on ICPSR, after?**

Title: Second-Generation Outcomes Of The Great Migration  
DOI: 10.1007/s13524-017-0625-8. Orig: F | LLM: T – **posted later on Dataverse**

Title: The Consequences Of Movement Office-Holding For Health Policy Implementation And Social Development In Urban Brazil  
DOI: 10.1093/sf/sox071. Orig: F | LLM: T – **not found online**

Title: Working Hours And Carbon Dioxide Emissions In The United States, 2007–2013. DOI: 10.1093/sf/soy014. Orig: F | LLM: T – **not found online**

Title: Missing Minorities? The Phases Of Irca Legislation And Relative Net Undercounts Of The 1990 Vis-À-Vis 2000 Decennial Census For Foreign-Born Cohorts. DOI: 10.1007/s13524-013-0212-6. Orig: F | LLM: T – **found 'supplementary material' but falsely coded**

Title: The Pragmatic American: Empirical Reality Or Methodological Artifact? DOI: 10.1111/1745-9125.12035. Orig: T | LLM: F – **shared elsewhere / emailed**

Title: Neighborhood Structural Characteristics, Individual-Level Attitudes, And Youths' Crime Reporting Intentions. DOI: 10.1111/j.1745-9125.2010.00212.x. Orig: T | LLM: F – **shared elsewhere / emailed**

Title: Rule Following And Discretion At Government's Frontlines: Continuity And Change During Organization Socialization. DOI: 10.1093/jopart/mup025  
Orig: F | LLM: T – **not found online**

Title: Solidaristic Unionism And Support For Redistribution In Contemporary Europe  
DOI: 10.1017/s0043887117000107. Orig: F | LLM: T – **found ESS data**

Title: Keeping The Public Purse: An Experiment In Windfalls, Taxes, And The Incentives To Restrain Government. DOI: 10.1017/s0003055413000415. Orig: T | LLM: F – **Dataverse later (2016)**

Title: The Electoral Implications Of Candidate Ambiguity. DOI: 10.1017/s0003055409090066  
Orig: F | LLM: T – **'available upon request'**

Title: Administrative Unit Proliferation. DOI: 10.1017/s0003055413000567. Orig: T | LLM: F – **shared elsewhere online/ emailed (no repository information found or attached to article)**

Title: It's Not Just What You Have, But Who You Know: Networks, Social Proximity To Elites, And Voting In State And Local Elections. DOI: 10.1017/s000305541600071x. Orig: T | LLM: F – **shared elsewhere / emailed**

Title: The Spatial Structure Of Party Competition: Party Dispersion Within A Finite Policy Space. DOI: 10.1017/s0007123409990172. Orig: T | LLM: F – **shared elsewhere / emailed**

Title: Compensation Or Constraint? How Different Dimensions Of Economic Globalization Affect Government Spending And Electoral Turnout. DOI: 10.1017/s0007123413000422. Orig: T | LLM: F – **shared elsewhere / emailed**

Title: Solidaristic Unionism And Support For Redistribution In Contemporary Europe. DOI: 10.1017/s0043887117000107. Orig: F | LLM: T – **'available upon request'**

Title: Precolonial Legacies And Institutional Congruence In Public Goods Delivery. DOI: 10.1017/s0043887117000363. Orig: T | LLM: F – **Dataverse, link hidden in Supplementary materials**

Title: Displacing Misinformation About Events: An Experimental Test Of Causal Corrections  
DOI: 10.1017/xps.2014.22. Orig: T | LLM: F – **shared elsewhere / emailed**



## Sociologists' Data Sharing Behaviors at Parity?

---

- In 2018, Political Scientists had way more data online than Sociologists
- In 2018, Political Scientists more often had a replication package ready and/or were more willing to respond to requests for sharing it
- Sociology studies published before 2018 rarely *had* data shared online (1%, but let's say less than 5% to be in an error margin)
- By 2026, this has jumped to 27% *for those same studies* (~moderate reliability)
  - Based on the original study, this is an underestimation (Maybe 40% is likely if we count personal websites or email communications)
  - Could be at parity with political science
  - Could have been a bad luck sample, although many of those studies simply do not share data after checking

# What about Robustness and Replicability?

---

## Robustness

- Sociology had <10 papers. Not enough for discipline-specific findings

## Replicability

- Sociology did as well as the others
- Paper-level
  - Sociology 51.5%
  - Economics 42.5%
  - Poli Sci 52.0%
- Claims-level
  - Sociology 45.0%
  - Economics 47.0%
  - Poli Sci 66.0%

# Why Wouldn't We Share Data in Sociology?

---

## Notes on Replication

Andrew Abbott  
*American Journal of Sociology*

Sociological Methods  
& Research  
Volume 36 Number 2  
November 2007 210-219  
© 2007 Sage Publications  
10.1177/0049124107306664  
<http://smr.sagepub.com>  
hosted at  
<http://online.sagepub.com>

This comment argues that although replication will and should gain ground in sociology, that process will be complicated by issues of ownership, mechanics, and security. Replicationism will also change the economy of peer review. Ironically, it could also reveal that sociologists have less agreement on methodological issues than we think.

**Keywords:** replication; data ownership; security; journals; peer review

Replication in the social sciences is a growing social movement, as those familiar with economics journals can attest. Jeremy Freese's piece (2007 [this issue]) invites sociologists to join that movement, providing a variety of rationales for doing so. In this comment, I shall not try

- argues lack of clarity or hidden code mistakes are *good*, can find “new” truths
- Argues that ‘replicationism’ will make us sterile and more like economics

## Replication, Replication

Gary King,<sup>1</sup> *Harvard University*

Political science is a community enterprise; the community of empirical political scientists needs access to the body of data necessary to replicate existing studies to understand, evaluate, and especially build on this work. Unfortunately, the norms we have in place now do not encourage, or in some cases even permit, this aim. Following are suggestions that would facilitate replication and are easy to implement—by teachers, students, dissertation writers, graduate programs, authors, reviewers, funding agencies, and journal and book editors.

### Problems in Empirical Political Science

As virtually every good methodology text explains, *the only way to*

How were the respondents selected? Who did the interviewing? What was the question order? How did you decide which informants to interview or villages to visit? How long did you spend in each community? Did you speak to people in their language or through an interpreter? Which version of the ICPSR file did you extract information from? How knowledgeable were the coders? How frequently did the coders agree? Exactly what codes were originally generated and what were all the recodes performed? Precisely which measure of unemployment was used? What were the exact rules used for conducting the content analysis? When did the time series begin and end? What countries were included in your study and how were they chosen? What statistical procedures were used? What method of numer-

- Wholesale support of replication as the necessary way forward

# A Lack of Open Science Norms in Sociology

---

- Sociology journals – wholesale lack of enforced transparency policies
- Translocational reproducibility – it is changing fast
- Fear of the ‘replication police’, data sharing practices
- Nothing to gain? Replication is not popular in sociology
  - Breznau (2022) ‘Does Sociology Need Open Science?’
    - Sociology does not do retractions,
    - <40 replications prior to 2020 in top journals
  - Philip Cohen (2026) ‘Update on ASA’
    - Recent volume of ASR, two-thirds had replication materials
    - 2020-2023 was closer to one-quarter

Thank you for your attention.



German Institute for  
Adult Education

Leibniz Centre for  
Lifelong Learning

[www.die-bonn.de](http://www.die-bonn.de)



Ministry of Culture and Science  
of the German State  
of North Rhine-Westphalia



With funding from the:



Federal Ministry  
of Research, Technology  
and Space

# References

---

- Breznau, Nate. 2021. "Does Sociology Need Open Science?" *Societies* 11(1):9. doi:[10.3390/soc11010009](https://doi.org/10.3390/soc11010009).
- Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, et al. 2025. "The Reliability of Replications: A Study in Computational Reproductions." *Royal Society Open Science* 12(3):241038. doi:[10.1098/rsos.241038](https://doi.org/10.1098/rsos.241038).
- Breznau, Nate, Eike Mark Rinke, Alexander Wuttke et al. 2022. "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty." *Proceedings of the National Academy of Sciences* 119(44):e2203150119. doi:[10.1073/pnas.2203150119](https://doi.org/10.1073/pnas.2203150119).
- Brodeur, Abel et al. 2026. "Reproducibility and Robustness of Economics and Political Science Research." *Nature* 652(8108):151–56. doi:[10.1038/s41586-026-10251-x](https://doi.org/10.1038/s41586-026-10251-x).
- Chang, Andrew, and Phillip Li. 2022. "Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say 'Often Not.'" *Critical Finance Review* 11(1):185–206. doi:[10.1561/104.00000053](https://doi.org/10.1561/104.00000053).
- Cohen, Philip N. 2025. "Another Update on American Sociological Review's Failure to Meet Current Social Science Standards." <https://familyinequality.wordpress.com/2025/12/10/another-update-on-american-sociological-reviews-failure-to-meet-current-social-science-standards/>.
- Eubank, Nicholas. 2016. "Lessons from a Decade of Replications at the Quarterly Journal of Political Science." *PS: Political Science & Politics* 49(2):273–76. doi:[10.1017/S1049096516000196](https://doi.org/10.1017/S1049096516000196).
- Galiani, S., Gertler, P., & Romero, M. (2018). Incentives for replication in economics. *Economic Inquiry*, 56(1), 458–472. <https://doi.org/10.1111/ecin.12502>
- Hardwicke, T. E., et al. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, 5(8), 180448. <https://doi.org/10.1098/rsos.180448>
- Janz, N. (2015). Leading journal verifies articles before publication – So far, all replications failed. Political Science Replication Blog.
- Krähmer, Daniel, Laura Schächtele, and Katrin Auspurg. 2026. "Code Sharing and Reproducibility in Survey-Based Social Research: Evidence from a Large-Scale Audit." *Royal Society Open Science* 13(3):251997. doi:10.1098/rsos.251997
- Liu, D. M., & Salganik, M. J. (2019). Successes and struggles with computational reproducibility: Lessons from the Fragile Families Challenge. *Socius*, 5. <https://doi.org/10.1177/2378023119849803>
- Stockemer, D., Koehler, S., & Lentz, T. (2018). Data access, transparency, and replication: New insights from the political behavior literature. *PS: Political Science & Politics*, 51(4), 799–803. <https://doi.org/10.1017/s1049096518000926>
- Trisovic, A., Lau, M. K., Pasquier, T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, 9, 60. <https://doi.org/10.1038/s41597-022-01143-6>