Article

Investigating the reproducibility of the social and behavioural sciences

https://doi.org/10.1038/s41586-026-10203-5

Received: 21 March 2025

Accepted: 27 January 2026

Check for updates

Published claims should be reproducible, yielding the same result when the same analysis is applied to the same data[1,2]. Here we assess reproducibility in a stratified random sample of 600 papers published from 2009 to 2018 in 62 journals spanning the social and behavioural sciences. The authors of 144 (24.0%, 95% confidence interval (CI) = 20.8–27.6%) papers made data available to assess reproducibility and, for 38 others, we obtained source data to reconstruct the dataset. We assessed 143 out of the 182 available datasets and found that 76.6 (53.6%, 95% CI = 45.8–60.7%) papers were rated as precisely reproducible and 105.0 (73.5%, 95% CI = 66.4–80.0%) were rated as at least approximately reproducible (within 15% of the original effects or within 0.05 of original *P* values) after inverse weighting each of the 551 claims by the number of claims per paper. We observed higher reproducibility for papers from political science and economics compared with other fields, for more recent papers compared with older papers and for papers from journals that require data sharing. Implementation of measures to verify that research is reproducible is needed to support trustworthiness in the complex enterprise of knowledge production[3,4].

Readers of quantitative research are sceptical of whether a research design justifies the authors' conclusions, raising questions of whether the chosen measures are valid assessments of the constructs of interest; whether the findings would be the same with a different analytical specification; and whether the findings will generalize to other circumstances. Scepticism identifies weaknesses, roots out errors and suggests alternative explanations to investigate. However, even sceptical readers will ordinarily assume that the quantitative analyses and outcomes are reported precisely.

Productive scholarly dialogue is difficult if readers of papers are left wondering whether the reported sample size is the same as the sample size in the dataset, whether the reported means reflect the actual means from the data or whether the reported model is the model that the authors used in their analysis. Ideally, readers should be able to assume that the described analysis, applied to the original data, consistently produces the reported outcomes. This Article investigates how close we are to this ideal.

Investigations in economics, finance, political science, cognitive science, psychology, social sciences, health, ecology and elsewhere suggest that reproducibility[1,2], defined as observing the same results when applying the same analysis to the same data, cannot be taken for granted[3–18]. Irreproducible outcomes can occur because of coding mistakes, transcription errors or faulty record keeping, many of which are unintentional and all of which are unwelcome.

Investigations of reproducibility hinge on the accessibility of the data and often the analytical code, because descriptions of analytical methods may be incomplete or difficult to translate back to code[10,19,20]. We assessed the availability of author-provided data to enable an independent test of reproducibility. 'Author-provided data' refers to data that authors made available through a website, repository or direct correspondence by email. This could have been raw or source data, prepared or derived data, or a combination. Data availability rates of well below 50%, and sometimes in single digits, have been reported in the fields of biomedicine, cancer biology, ecology, business and economics, and across the social sciences more generally[9,12,21–24].

As part of the DARPA-funded Systemizing Confidence in Open Research and Evidence (SCORE) program[25], we conducted a systematic investigation of reproducibility in the social and behavioural sciences.

Data and code availability

We assessed data availability for 600 papers from a stratified random sample published from 2009 to 2018 in 62 journals across the social and behavioural sciences (Table 1 and Supplementary Table 4). For each paper, we searched for data collected or prepared by the original author(s) for the analyses reported in the paper. The author-provided data were considered available if they were publicly available or made accessible by the authors on request. We examined the paper's main text and supplementary materials for the data and links to repositories, and we recorded whether authors explicitly stated that some or all of the original data sources were restricted or unable to be shared for ethical or legal reasons. Restricted data were counted as not available (13 cases observed, all from economics). We did not systematically document whether source data were available if the authors did not make them available themselves. We conducted a similar search for analytical code that implemented the analyses reported in the paper.

We obtained both data and code for 120 articles (20.0%, 95% CI = 17.0–23.4%), just data for 24 (4.0%, 95% CI = 2.7–5.9%), just code for 26 (4.3%, 95% CI = 3.0–6.3%) and neither for 430 (71.7%, 95% CI = 67.9–75.1%). Thus, data were available for 144 papers (24.0%, 95% CI = 20.8–27.6%) and unavailable for 456 papers (76.0%, 95% CI = 72.4–79.2%).

Data and code availability by year

Previous investigations suggest that data and code availability is higher for more recent papers—perhaps owing to poor archival practices that lead to data loss over time or to improvements in data sharing standards over time[12,26]. Our sample spans a 10-year time period of widespread

A list of authors and their affiliations appears at the end of the paper.

Table 1 | The 62 journals included in the sample for selecting papers and claims

Business	**Education**	**Psychology**
Academy of Management Journal	*American Educational Research Journal*	*Child Development*
Journal of Business Research	*Computers and Education*	*Clinical Psychological Science*
Journal of Management	*Contemporary Educational Psychology*	*Cognition*
Leadership Quarterly	*Educational Researcher*	*European Journal of Personality*
Management Science	*Exceptional Children*	*Evolution and Human Behavior*
Organization Science	*Journal of Educational Psychology*	*Journal of Applied Psychology*
Journal of the Academy of Marketing Science	*Learning and Instruction*	*Journal of Consulting and Clinical Psychology*
Journal of Consumer Research		*Journal of Environmental Psychology*
Journal of Marketing		*Journal of Experimental Psychology: General*
Journal of Marketing Research		*Journal of Experimental Social Psychology*
Journal of Organizational Behavior		*Journal of Personality and Social Psychology*
Org. Behavior and Human Decision Processes		*Psychological Science*
		Health Psychology
		Psychological Medicine
		Social Science and Medicine
Economics	**Political science**	**Sociology**
American Economic Journal: Applied Economics	*American Journal of Political Science*	*American Journal of Sociology*
American Economic Review	*American Political Science Review*	*American Sociological Review*
Econometrica	*British Journal of Political Science*	*Demography*
Experimental Economics	*Comparative Political Studies*	*European Sociological Review*
Journal of Finance	*Journal of Conflict Resolution*	*Journal of Marriage and Family*
Journal of Financial Economics	*Journal of Experimental Political Science*	*Social Forces*
Journal of Labor Economics	*World Politics*	*Criminology*
Journal of Political Economy	*Journal of Public Admin. Research and Theory*	*Law and Human Behavior*
Quarterly Journal of Economics	*Public Administration Review*	
Review of Financial Studies		
World Development		

For primary reporting, economics and finance were combined as economics; sociology and criminology were combined as sociology; management, marketing and organizational behaviour were combined as business; psychology and health were combined as psychology; and political science and public administration were combined as political science. Outcomes are reported separately by subfield in the Supplementary Information.

discussion of reproducibility and changes to journals' reproducibility policies, offering a window for observing variation due to both of these factors. We replicated the association between year of publication and availability of data and code, with more recent papers having higher rates of sharing data, code or both (Fig. 1). This is also reflected in modest positive correlations between the year and the percentage of papers with data available (ρ = 0.16, 95% CI = 0.08–0.24), code available (ρ = 0.17, 95% CI = 0.09–0.24) or both available (ρ = 0.16, 95% CI = 0.08–0.24).

Conditional on data or code being available, we did not observe clear evidence of greater data and code availability for more recent papers. Overall, 120 out of 170 papers (70.6%, 95% CI = 63.3–76.9%) with either data or code available had both data and code available, and the correlation with year of publication was ρ = 0.07 (95% CI = −0.08–0.22). In summary, considering papers for which some sharing occurred, the comprehensiveness of sharing was not significantly higher for more recent papers.

Data and code availability by field

Papers from the 62 included journals were aggregated into six field categories for expository purposes: business (including marketing, management and organizational behaviour), economics (including finance), education, political science (including public administration), psychology (including health) and sociology (including criminology). Figure 2 illustrates that political science (46 out of 85 papers, 54.1%, 95% CI = 43.6–64.3%) and economics (38 out of 102 papers, 37.3%, 95% CI = 28.5–46.9%) had higher data availability rates compared with the other fields (combined: 14.5%, 95% CI = 11.5–18.3%), and education had the lowest (2.9%, 95% CI = 0.8–10.0%). The figure also illustrates that political science and economics have similar data availability rates, if the 13 cases of restricted data in economics are considered available, in principle. Extended Data Fig. 1 shows even higher data availability in political science (69.2%, 95% CI = 57.2–79.1%) and economics (48.6%, 95% CI = 37.4–59.9%) after they are separated from public administration (5.0%, 95% CI = 0.3–23.6%) and finance (10.0%, 95% CI = 3.5–25.6%), respectively (Extended Data Figs. 2 and 3).

Assessing reproducibility

Whereas we evaluated data availability only at the paper level, reproducibility was also assessed for individual claims within papers. For most papers, we extracted and evaluated a single key claim but, for a subset of papers, we extracted multiple claims per paper[27]. In total, in 59 out of ~~145 (40.7%)~~ papers, more than 1 claim was assessed for reproducibility (mean claims per paper = 3.9, s.d. = 5.9, range = 1–37). In total, 551 claims from 143 papers were assessed for reproducibility.

The reproducibility of multiple claims within a paper is potentially statistically dependent, because they arise from the same project and authors. We assessed reproducibility at (1) the claim level and (2) the paper level by weighting claims-level data to the paper level (for example, if there were four claims in a paper, each was weighted to be equivalent to 0.25 observations) and clustering to account for interclass correlation among claims. As such, reproducibility for papers can be fractional based on the outcomes of claims within the same paper. We report paper-level outcomes in the main text and claim-level outcomes in Extended Data Figs. 4–6.

Representativeness

As assessing reproducibility requires access to data, which varies by field and time, the subset of papers tested for reproducibility may not be fully representative of the broader sample. Table 2 shows, by field, how representativeness shifted at successive stages of the research process[27]. The initial stages of selecting papers and identifying claims maintained representativeness by field (Supplementary Tables 3 and 5).

The primary determinant of inclusion in the reproducibility tests was successful acquisition of author-generated data (n = 144). We also included 38 papers for which author data were unavailable but source data were obtainable to reconstruct the datasets. Together, these constitute the papers with source or author data available in Table 2.

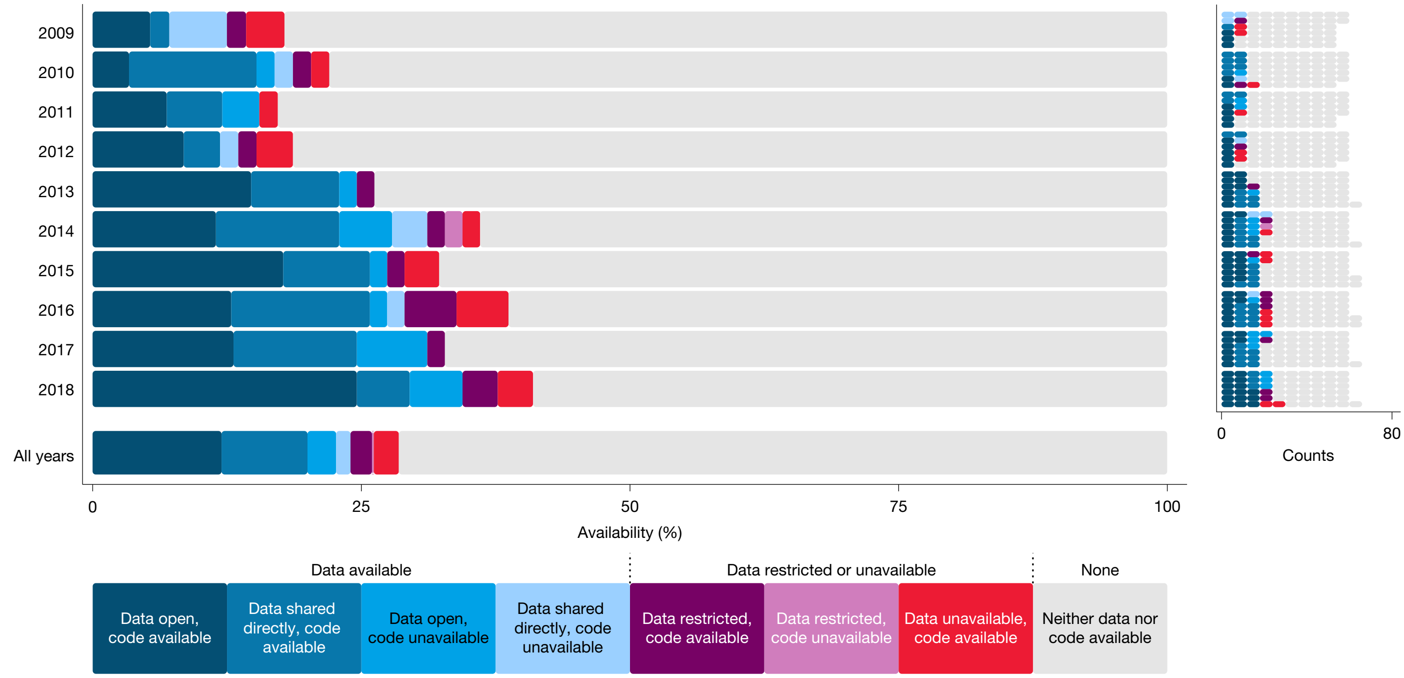


Fig. 1 | Data and code availability by year of publication. Left, data and code availability as a percentage of papers. Right, raw counts of papers with author-provided data and code available and not available. Restricted data (purple) did not count as available data, but might be accessible in principle.

Relative to the broader sample, political science and economics became a larger share of the tested sample; sociology was largely unchanged; and the remaining fields decreased in share. At the claim level, economics accounts for a higher proportion because its papers contained more reanalysed claims per paper. An analogous assessment of representativeness by publication year is shown in Supplementary Table 3.

Observed reproducibility

Analysts were matched with data to reanalyse and followed a structured protocol. Reproducibility was investigated with three possible outcomes: precise reproducibility, approximate reproducibility and not reproduced. Precise reproducibility was achieved if all of the statistical outcomes of the reproduction were the same as originally reported. This could include, for example, the sample size, focal regression coefficient, test statistic, effect size and P value for a single claim. Approximate reproducibility was defined a priori as achieved if at least one statistical outcome was not precisely reproduced and all outcomes for a claim were reproduced within ±15% of what was originally reported and, for P values, a difference of no more than 0.05. If any of the statistical outcomes were neither precisely nor approximately reproduced, then the claim was coded as not reproduced.

Whereas 146 papers and 555 claims had at least one outcome reproduction attempt, 4 claims had none of our eligible statistical outcomes and were not counted for the quantitative assessment of reproducibility. Claims were reweighted to the paper level after we excluded these claims. As such, 143 papers comprising 551 claims were assessed for reproducibility. For five papers, an outcome reproduction attempt

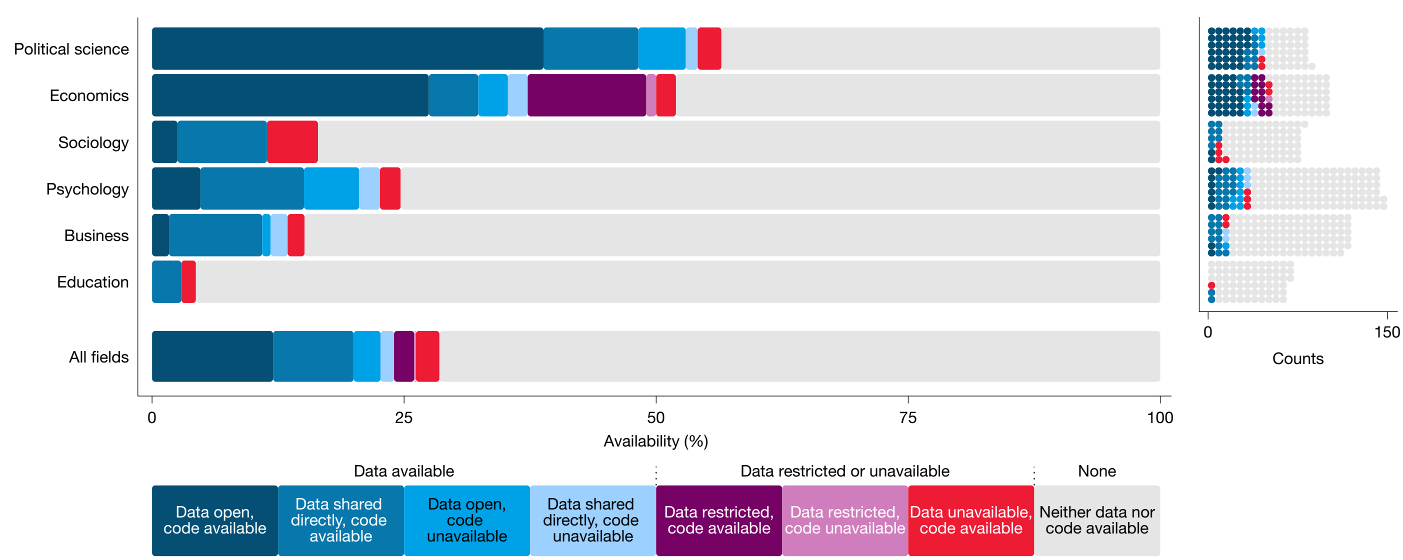


Fig. 2 | Data and code availability by field. Left, data and code availability as a percentage of papers. Right, the raw counts of papers with data and code available and not available. Restricted data (purple) did not count as available data, but might be accessible in principle.

Table 2 | Number of papers at each stage of the selection process and number and percentage of papers and claims reproduced by field

	Business	Economics	Education	Political science	Psychology	Sociology	Total
Papers with claims (*n* (%))	591 (19.7)	520 (17.3)	342 (11.4)	424 (14.1)	727 (24.2)	396 (13.2)	3,000 (100)
Papers eligible for reproduction (*n* (%))	119 (19.8)	102 (17.0)	69 (11.5)	85 (14.2)	146 (24.3)	79 (13.2)	600 (100)
Papers with multiple claims (*n* (%))	38 (19.0)	33 (16.5)	23 (11.5)	32 (16.0)	49 (24.5)	25 (12.5)	200 (100)
Papers with single claim (*n* (%))	81 (20.2)	69 (17.2)	46 (11.5)	53 (13.2)	97 (24.2)	54 (13.5)	400 (100)
Papers with source or author data available (*n* (%))	17 (9.3)	41 (22.5)	10 (5.5)	50 (27.5)	41 (22.5)	23 (12.6)	182 (100)
Papers with at least one claim reproduction started (*n* (%))	15 (9.1)	38 (23.0)	11 (6.7)	46 (27.9)	31 (18.8)	24 (14.5)	165 (100)
Papers with at least one claim reproduction completed (*n* (%))	14 (9.6)	31 (21.2)	9 (6.2)	43 (29.5)	28 (19.2)	21 (14.4)	146 (100)
Total reproductions of claims (*n* (%))	46 (7.4)	171 (27.6)	23 (3.7)	199 (32.1)	121 (19.5)	60 (9.7)	620 (100)
Reproductions of unique claims (*n* (%))	40 (7.2)	160 (28.8)	23 (4.1)	177 (31.9)	102 (18.4)	53 (9.5)	555 (100)

began, but the analysts determined that the material that they had was not sufficient to assess reproducibility. This could occur if the provided data were incomplete or otherwise compromised for conducting a reproduction or if the provided code were not usable or adaptable. These were counted as reproducibility failures.

Of the 143 papers that were assessed, we observed approximate or precise reproducibility for 105.0 papers (73.5%, 95% CI = 66.4–80.0%) and precise reproducibility for 76.6 papers (53.6%, 95% CI = 45.8–60.7%).

Figure 3 shows the reproducibility results separately for different circumstances of conducting the reproduction. When code and data were available, we attempted to execute the original code or adapt it if necessary. We observed approximate or precise reproducibility for 73.8 of the 81.2 papers (90.9%, 95% CI = 84.5–96.6%) and precise reproducibility for 62.7 of the 81.2 papers (77.2%, 95% CI = 68.9–84.7%). For 52.4 (64.5%, 95% CI = 55.9–74.2%) of these papers, we were able to reproduce the findings with minimal effort other than executing the code on the data, a standard known as push-button reproducibility[28].

When only data were available, we attempted to reproduce the findings by generating new code following the analyses described in the paper. Of these, we observed approximate or precise reproducibility for 16.1 out of the 22.1 papers (72.9%, 95% CI = 54.2–88.2%) and precise reproducibility for 9.5 of the 22.1 papers (43.2%, 95% CI = 25.5–61.4%).

When author-provided data were unavailable, but source data were available, we attempted to reproduce the findings by preparing the data and generating new code. Of these, we observed approximate or precise reproducibility for 15.1 out of the 39.7 papers (38.1%, 95% CI = 24.0–53.3%) and precise reproducibility for 4.4 of the 39.7 papers (11.0%, 95% CI = 3.8–20.2%). In summary, reproducibility rates were comparatively high when data and code were both available, and comparatively low when we needed to reconstruct the data and code[10].

In addition to the empirically defined reproducibility criteria, we asked analysts to provide their subjective assessment of whether they successfully reproduced each claim. This included papers and claims that did not have eligible statistical outcomes for our quantitative evaluation. Excluding missing or undetermined cases, analysts reported successful reproductions of 83 out of 132 papers (62.9%, 95% CI = 54.4–70.6%) and 433 out of 535 claims (80.9%, 95% CI = 77.4–84.0%).

Reproducibility by year

Figure 4 presents reproducibility by year. The number of reproduction attempts per year is quite small. Considering only papers with an attempt, the prevalence of precise reproducibility (Spearman's $\rho = -0.051$, 95% CI = −0.207–0.138) and the prevalence of approximate and precise reproducibility (Spearman's $\rho = -0.010$, 95% CI = −0.204–0.211) were not significantly associated with time.

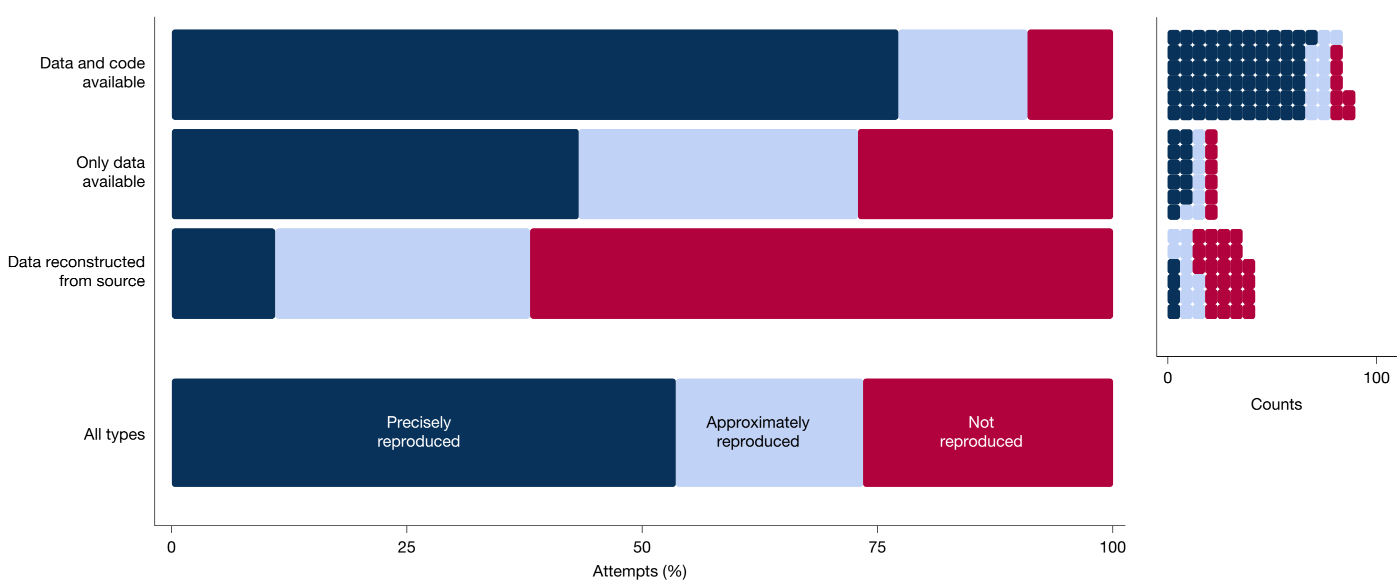


Fig. 3 | Reproducibility by data and code availability. Reproducibility as a percentage of attempts (left) and reproducibility as counts (right).

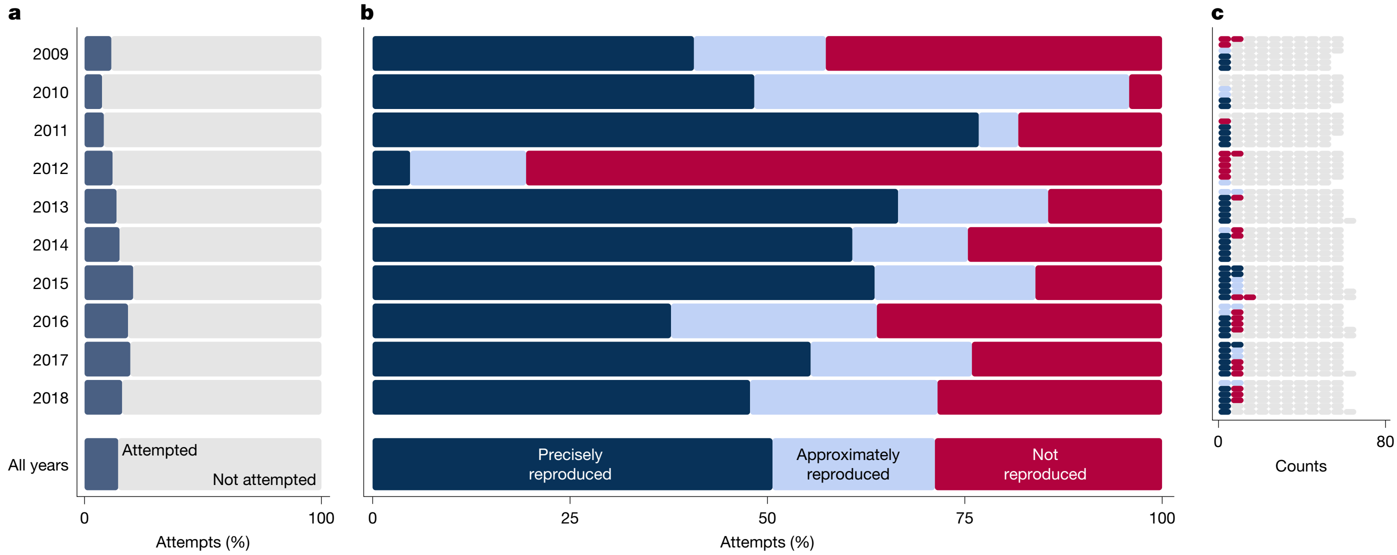


Fig. 4 | Reproducibility by year of publication. a, The proportion of reproduction attempts from the sample of papers. **b**, Reproducibility as a percentage of the attempts. **c**, Reproducibility as counts compared with the sample of papers. Note that papers with multiple claims could be partly reproducible, but the colour coding of the dots showing paper counts in **c** is rounded to the nearest paper.

Reproducibility by field

Figure 5 presents reproducibility by field (Extended Data Figs. 7–10). Political science and economics had much higher rates of reproduction attempts than other fields owing to greater data availability. Considering only papers with a reproduction attempt, we observed approximate or precise reproducibility for 35.2 out of 41.9 (84.0%, 95% CI = 72.7–93.3%) political science papers and 23.9 out of 31.0 (77.2%, 95% CI = 63.6–90.2%) economics papers. We observed precise reproducibility for 27.6 out of 41.9 (65.9%, 95% CI = 51.7–79.3%) political science papers and 22.2 out of 31.0 (71.6%, 95% CI = 56.4–86.7%) economics papers. Combining the data across the other four fields, we observed approximate or precise reproducibility for 45.8 out of 70.0 (65.4%, 95% CI = 54.9–76.2%) papers and precise reproducibility for 26.7 out of 70.0 (38.2%, 95% CI = 28.1–48.3%) papers.

Exploration of journal policies

We conducted a follow-up exploratory investigation of the relationship between journal policies and reproducibility (Supplementary Fig. 11 and Supplementary Tables 6–13). Data availability was observed more consistently for papers published in journals that require data sharing (87.5%), data and code sharing (66.2%), or data and code sharing and reproducibility checks (100.0%) than for papers published in journals with none of those policies (16.0%; $\chi^2_3 = 100.3$, $P < 0.001$). Conditional on attempting a reproduction, precise reproducibility was observed more frequently for papers published in journals that require data sharing (70.5%), data and code sharing (75.7%), or data and code sharing and reproducibility checks (65.0%) than for papers published in journals with none of those policies (40.7%), although the evidence was only suggestive ($\chi^2_6 = 16.2$, $P = 0.013$). Similar analyses at the claim level were consistent and statistically significant ($\chi^2_6 = 46.3$, $P < 0.001$). Analyses using an ordered logistic regression model at both the paper and claim levels suggested that the association between journal policy and reproducibility was primarily observed in distinguishing papers and claims that were not reproduced from papers and claims that were approximately and precisely reproduced.

The prevalence of journal policies requiring data sharing, code sharing and reproducibility checks has increased over time (Fig. 6). Reproducibility checks refer to the journal using an internal process to assess reproducibility before publication. In 2018, the last year of our sample of papers, 27.4% of journals had a data sharing requirement, 22.6% had a code sharing requirement and 3.2% conducted reproducibility checks. In total, 77.8% of journals from economics and political science in our sample and 6.8% of journals from other fields had at least one of these policies. As of mid-2025, rates had increased to 51.6% of journals having a data sharing requirement, 46.8% having a code sharing requirement and 19.4% conducting reproducibility checks. In total, 94.4% of journals from economics and political science and 43.2% of journals from other fields had at least one of the policies.

Discussion

A basic assumption of quantitative research is that repeating the same analysis on the same data will produce the same result as the original report; that is to say, that the reported result is reproducible. The most substantial barrier to observing reproducibility in our sample of social and behavioural science papers was the unavailability of author data, preventing reproduction attempts. When reproductions could be attempted, availability of data and code was associated with greater but imperfect reproducibility compared with only data availability. Attempting to reproduce findings from source data had a lower success rate. Political science and economics papers were more likely to have data available and reproduce successfully than those in other fields, and exploratory evidence introduces the hypothesis that this could be due partly to the presence of journal policies requiring the sharing of data or code or reproducibility checks before publication. Looking forward, the proportion of journals requiring adherence to these policies has increased since 2018—the year in which the most recent papers in our sample were published. These findings provide several insights about reproducibility in the social and behavioural sciences.

Data availability

A lack of data availability does not mean that the outcomes cannot be trusted. It is possible that the results would be perfectly reproducible if the data were available. The primary consequence of this is uncertainty: readers do not know whether the results are reported precisely.

Our criterion was the availability of author-provided data for conducting the analyses reported in the paper. It is likely that more data and code could have been accessed if we had adopted more assertive methods to obtain it. This includes pursuing restricted data that are ostensibly available by meeting the requirements to gain access.

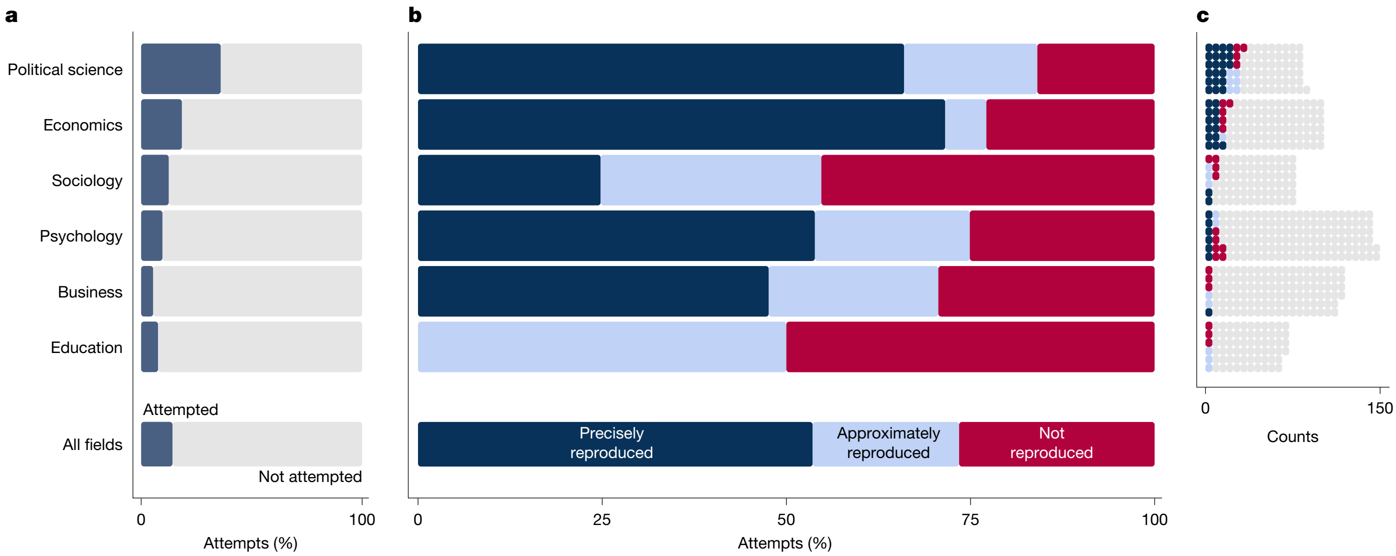


Fig. 5 | Reproducibility by field. a, The proportion of reproduction attempts from the sample of papers. **b**, Reproducibility as a percentage of the attempts. **c**, Reproducibility as counts compared with the sample of papers. Note that (1) papers with multiple claims could be partly reproducible, but the colour coding of the dots showing paper counts in **c** is rounded to the nearest paper; and (2) the weighting scheme in **c** includes claims without reproduction attempts, so the proportions will not exactly match the middle bars.

For example, data-use agreements for confidential data have been used at AEA journals for verifying reproducibility[29].

We could have relaxed the definition of what counted as data availability beyond author-provided datasets, such as including occasions for which source data could be found and obtained, although we observed much less reproducibility in such cases. There are often several data-management steps between source data (or raw data) and a dataset that is prepared for inferential analyses (or processed data) that will be reported in the paper. Sometimes these steps are represented in available analytical code, and sometimes they are not. There may be a trade-off between data availability standards and reproducibility. Greater leniency on what counts as sufficient data sharing may be associated with greater failures in reproducing the outcomes. More complete and precise documentation of the data preparation and analysis pipeline improves transparency and, possibly, reproducibility[20].

The estimated reproducibility differs considerably depending on whether papers with no data available are included or ignored—for example, 17.6% versus 73% for approximate or precise reproducibility. Which is a more appropriate reproducibility estimate depends on one's perspective. If the question of interest is whether the outcomes can be verified, then the low estimate reflects the percentage of outcomes that we were able to reproduce independently given the amount of effort we invested in gaining access to data and conducting reanalysis. If the question of interest is whether outcomes are reported precisely in papers, then the higher reproducibility estimate might be closer to reality. Presumably, some of the papers with unavailable data would have reproduced successfully if the data could have been obtained.

Reproducibility failures

A reproducibility failure does not mean that the finding is wrong. A reproducibility test can fail because the data used are not identical to the original data; the code or computational environment used to execute the code used is not aligned with the original analysis; the reproduction analyst makes an error or did not spend enough time troubleshooting; or the original description of the data preparation and analysis was incomplete or inaccurate. If those are the sole reasons why the reproduction test failed, then the original outcome may have been reported precisely despite the independent failure to reproduce them. Previous evidence suggests imperfect consistency across reproducibility analysts, and perhaps more so when working from source data rather than author-provided datasets, suggesting that this has a meaningful role[9,30]. Even so, a failure in this context creates undesirable uncertainty regarding the credibility of the original outcomes.

We also observed a sizable number of approximately correct reproductions. We defined approximately correct statistically—within 15% of the original effects or within 0.05 of the P value. Previous investigations in which several analysts investigate the same question with the same data have found substantial variation in data preparation and analysis decisions that may not be reported clearly[31–33]. Without code provided, reproduction attempts may involve making reasonable inferences about how the data were prepared and analysed based on what is written in the paper. Furthermore, even when code is available, other research suggests that reproducibility is higher when researchers have better coding skills, investigate simpler research questions and have simpler code[9]. With code provided, reproduction attempts may vary in complexity, increasing the odds of approximate rather than precise success. As such, even though all findings should be precisely reproducible in principle, there are practical challenges in conducting reproductions that may reduce precision without indicting the original finding[34].

Reproducibility successes

Reproducibility success does not mean that the finding is correct. Reproducibility means that the results are reported precisely. Precisely reported results can be wrong because the analysis strategy is invalid, there are coding errors in the data, the research design is confounded, the result is not robust to reasonable alternative analytical decisions or the researcher selectively reported positive results from many analyses, inflating the likelihood of exaggerated findings[35–39]. Reproducibility is a baseline assessment of credibility for quantitative findings[40].

Differences across fields

Papers from political science and economics were more likely to have data available and achieve reproducibility compared with papers from other fields. A follow-up exploratory investigation suggested an association between reproducibility success and journal policies requiring data sharing, code sharing and reproducibility checks. Economics and political science journals were more likely to have such policies. Moreover, transparency policies have become more popular over time across

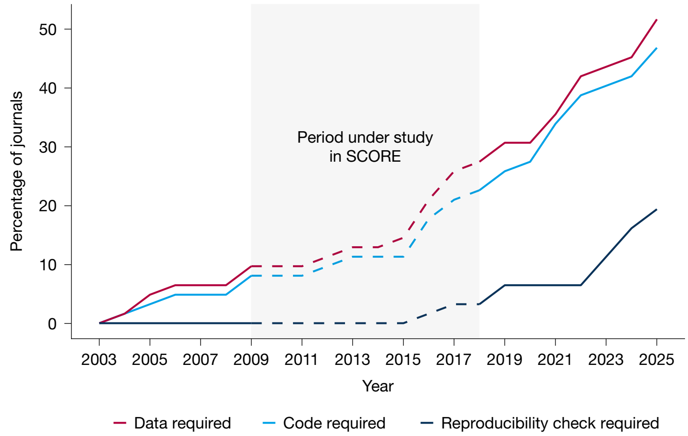


Fig. 6 | The percentage of 62 journals with data sharing, code sharing and reproducibility check requirements from 2003 to 2025. Policies for all 62 journals are included, although three had no reproduction attempts (*Journal of Finance*, *Journal of Public Administration Research and Theory* and *Law and Human Behavior*).

our sample of social and behavioural science journals. If transparency policies have a causal impact on reproducibility, then reproducibility success may be higher in a replication with the same journals using a more recent sample of papers. Future investigations into the causes of reproducibility could also assess the role of social norms, training, tools used in data preparation and analysis, and potential variation across research methodologies. There could be interactions between causes. For example, a transparency policy could cause an increase in data availability and a decrease in reproducibility if implementing the policy is highly burdensome: available data and code might be unusable by independent researchers, thereby harming reproducibility attempts.

Constraints on generalizability

We conducted reproducibility tests on a stratified sample of papers published from 2009 to 2018 from 62 journals in the social and behavioural sciences. Included papers had to have a quantitative outcome associated with a primary claim in the abstract of the paper. Selection of the 62 journals followed a principled approach that was applied consistently across disciplinary boundaries. Nevertheless, the overall and field-specific rates may differ with a different sample of journals. Likewise, the exploratory findings that reproducibility rates vary by time and transparency policies imply that the observed outcomes may be different during other time periods. The papers subjected to data and code availability assessment remained representative of the sample, but we did not attempt to access a small number of datasets that were reported as restricted but could, in principle, be obtained. The papers subjected to reproducibility assessment were skewed because a test could be conducted only if data were available. The extent to which this affects the generalizability is unknown. In every field for which a reproducibility study has been conducted, both data availability and reproducibility have fallen short of perfection[3,7,11,21]. Given that not all papers could be assessed for reproducibility, our reproducibility estimate may not generalize to our full sample or to the social and behavioural sciences generally. Even so, this evidence does suggest that reproducibility practices can improve in all fields investigated.

Limitations to data accessibility

Even if all findings are reported precisely, there are occasions in which reproducibility will not be easily verified because of barriers to data access (for example, a study on the role of method provenance in responsible FAIR reuse)[41]. Principal challenges recognized in open science policies and principles are privacy and proprietary data concerns[42]. In this project, we considered only data made available directly. Some data cannot be publicly shared because they contain personally identifiable or other sensitive information, or are proprietary data belonging to a firm or other private entity. There are a variety of solutions available to advance confidence in reproducibility even under these circumstances, although sometimes with substantial cost[43,44]. For example, some datasets can be anonymized to be publicly shareable for the purposes of demonstrating reproducibility of key findings[45]. Other datasets may not be anonymized, but can be archived and reanalysed under protected conditions through a variety of data centres with appropriate security and ethical oversight[46]. For proprietary data, authors can describe the process by which they obtained permission to use the data so that an independent researcher could follow the same steps for verification purposes. In some cases, the raw data may not be shareable, but the code and derived data could at least enable verification of the analysis and reporting workflow. Synthetic datasets can be created that reproduce the statistical outcomes without violating confidentiality concerns[47–49]. These solutions demonstrate that the aims of open science and security are not inevitably in opposition. Innovative approaches to data access can maintain proper controls for privacy and security and still provide pathways to accessibility and confidence in reproducibility.

Conclusion

Even the most experienced researchers will make errors in data management, analysis, recordkeeping and transcription. Implementing measures to verify that research is reproducible is not a statement that researchers are untrustworthy, but a recognition that high standards for quality control are needed because even the most diligent researchers will sometimes be unable to detect and correct mistakes.

A credo of the open scholarship movement is "as open as possible, as closed as necessary"[50,51]. Transparency and sharing enable independent observers to interrogate and verify the basis of research claims. Limitations in transparency and sharing may be inevitable in some cases, and deliberate efforts to maximize verifiability in those circumstances will benefit the trustworthiness of the research. Reproducibility failures add unnecessary uncertainty to the complex enterprise of knowledge production.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-10203-5.

1. Dreber, A. & Johannesson, M. A framework for evaluating reproducibility and replicability in economics. *Econ. Inq.* **63**, 338–356 (2025).
2. National Academies of Sciences, Engineering and Medicine. *Reproducibility and Replicability in Science* (The National Academies Press, 2019).
3. Culina, A., van den Berg, I., Evans, S. & Sánchez-Tójar, A. Low availability of code in ecology: a call for urgent action. *PLoS Biol.* **18**, e3000763 (2020).
4. Stodden, V., Seiler, J. & Ma, Z. An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc. Natl Acad. Sci. USA* **115**, 2584–2589 (2018).
5. Chang, A. C. & Li, P. Is economics research replicable? Sixty published papers from thirteen journals say "often not". *Crit. Finance Rev.* **11**, 185–206 (2022).
6. McCullough, B. D., McGeary, K. A. & Harrison, T. D. Do economics journal archives promote replicable research? *Can. J. Econ.* **41**, 1406–1420 (2008).
7. Vilhuber, L. Reproducibility and replicability in economics. *Harv. Data Sci. Rev.* **2**, 4 (2020).
8. Brodeur, A. et al. *Mass Reproducibility and Replicability: A New Hope* No. 107, I4R Discussion Paper Series (The Institute for Replication, 2024).
9. Pérignon, C. et al. Computational Reproducibility in finance: evidence from 1,000 tests. *Rev. Financ. Stud.* **37**, 3558–3593 (2024).
10. Laurinavichyute, A., Yadav, H. & Vasishth, S. Share the code, not just the data: a case study of the reproducibility of articles published in the *Journal of Memory and Language* under the open data policy. *J. Mem. Lang.* **125**, 104332 (2022).
11. Wang, S. V., Sreedhara, S. K. & Schneeweiss, S. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. *Nat. Commun.* **13**, 5126 (2022).

12. Minocher, R., Atmaca, S., Bavero, C., McElreath, R. & Beheim, B. Estimating the reproducibility of social learning research published between 1955 and 2018. *R. Soc. Open Sci.* **8**, 210450 (2021).
13. Hardwicke, T. E. et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *R. Soc. Open Sci.* **5**, 180448 (2018).
14. Eubank, N. Lessons from a decade of replications at the *Quarterly Journal of Political Science*. *PS Polit. Sci. Polit.* **49**, 273–276 (2016).
15. Trisovic, A., Lau, M. K., Pasquier, T. & Crosas, M. A large-scale study on research code quality and execution. *Sci. Data* **9**, 60 (2022).
16. Roche, D. G., Kruuk, L. E. B., Lanfear, R. & Binning, S. A. Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol.* **13**, e1002295 (2015).
17. Roche, D. G. et al. Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution. *Proc. R. Soc. B Biol. Sci.* **289**, 20212780 (2022).
18. Stockemer, D., Koehler, S. & Lentz, T. Data access, transparency, and replication: new insights from the political behavior literature. *PS Polit. Sci. Polit.* **51**, 799–803 (2018).
19. Brodeur, A. et al. Promoting reproducibility and replicability in political science. *Res. Polit.* https://doi.org/10.1177/20531680241233439 (2024).
20. Breznau, N. et al. The reliability of replications: a study in computational reproductions. *R. Soc. Open Sci.* **12**, 241038 (2025).
21. Errington, T. M., Denis, A., Perfito, N., Iorns, E. & Nosek, B. A. Challenges for assessing replicability in preclinical cancer biology. *eLife* **10**, e67995 (2021).
22. Gabelica, M., Bojčić, R. & Puljak, L. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *J. Clin. Epidemiol.* **150**, 33–41 (2022).
23. Khan, N., Thelwall, M. & Kousha, K. Data sharing and reuse practices: disciplinary differences and improvements needed. *Online Inf. Rev.* **47**, 1036–1064 (2023).
24. Tedersoo, L. et al. Data sharing practices and data availability upon request differ across scientific disciplines. *Sci. Data* **8**, 192 (2021).
25. Alipourfard, N. et al. Systematizing confidence in open research and evidence (SCORE). Preprint at *OSF* https://doi.org/10.31235/osf.io/46mnb (2021).
26. Vines, T. H. et al. The availability of research data declines rapidly with article age. *Curr. Biol.* **24**, 94–97 (2014).
27. Abatayo, A. L. et al. Assessments of credibility in the social and behavioural sciences. Preprint at *MetaArXiv* https://doi.org/10.31222/osf.io/7u58q_v1 (2026).
28. Wood, B. D. K., Müller, R. & Brown, A. N. Push button replication: is impact evaluation evidence for international development verifiable? *PLoS ONE* **13**, e0209416 (2018).
29. Vilhuber, L. & Cavanagh, J. Report of the AEA data editor. *AEA Pap. Proc.* **115**, 944–957 (2025).
30. Fišar, M., Greiner, B., Huber, C., Katok, E. & Ozkes, A. I. Reproducibility in management science. *Manag. Sci.* **70**, 1343–1356 (2024).
31. Huntington-Klein, N. et al. The influence of hidden researcher decisions in applied microeconomics. *Econ. Inq.* **59**, 944–960 (2021).
32. Silberzahn, R. et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).
33. Aczel, B. et al. Investigating the analytical robustness of the social and behavioural sciences. *Nature* https://doi.org/10.1038/s41586-025-09844-9 (2026).
34. Clemens, M. A. The meaning of failed replications: a review and proposal. *J. Econ. Surv.* **31**, 326–342 (2017).
35. Brown, A. W., Kaiser, K. A. & Allison, D. B. Issues with data and analyses: errors, underlying themes, and potential solutions. *Proc. Natl Acad. Sci. USA* **115**, 2563–2570 (2018).
36. Gelman, A. & Carlin, J. Beyond power calculations: assessing type S (Sign) and type M (magnitude) errors. *Perspect. Psychol. Sci.* **9**, 641–651 (2014).
37. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
38. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
39. Cummins, J. The threat of analytic flexibility in using large language models to simulate human data: a call to attention. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2509.13397 (2025).
40. Nuijten, M. B., Bakker, M., Maassen, E. & Wicherts, J. M. Verify original results through reanalysis before replicating. *Behav. Brain Sci.* **41**, e143 (2018).
41. Weissgerber, T. L. et al. Understanding the provenance and quality of methods is essential for responsible reuse of FAIR data. *Nat. Med.* **30**, 1220–1221 (2024).
42. Nosek, B. A. et al. Promoting an open research culture. *Science* **348**, 1422–1425 (2015).
43. Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R. & Debonnel, E. Certify reproducibility with confidential data. *Science* **365**, 127–128 (2019).
44. Vilhuber, L. Report of the AEA Data Editor. *AEA Pap. Proc.* **114**, 878–890 (2024).
45. Morehouse, K. N., Kurdi, B. & Nosek, B. A. Responsible data sharing: identifying and remedying possible re-identification of human participants. *Am. Psychol.* **80**, 928–941 (2025).
46. Nab, L. et al. OpenSAFELY: a platform for analysing electronic health records designed for reproducible research. *Pharmacoepidemiol. Drug Saf.* **33**, e5815 (2024).
47. Quintana, D. S. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *eLife* **9**, e53275 (2020).
48. Boedihardjo, M., Strohmer, T. & Vershynin, R. Covariance's loss is privacy's gain: computationally efficient, private and accurate synthetic data. *Found. Comput. Math.* **24**, 179–226 (2024).
49. Khan, A. R. & Kabir, E. Resampling methods for generating continuous multivariate synthetic data for disclosure control. *J. Data Inf. Manag.* **3**, 225–235 (2021).
50. *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020* (European Commission, 2016).
51. Landi, A. et al. The "A" of FAIR—as open as possible, as closed as necessary. *Data Intell.* **2**, 47–55 (2020).

Olivia Miske[1], Anna Lou Abatayo[2], Mason Daley[1], Mirka Dirzo[1], Nicholas Fox[1], Noah Haber[1], Krystal M. Hahn[1], Melissa Kline Struhl[3], Brinna Mawhinney[1], Priya Silverstein[4,5], Theresa Stankov[1], Andrew H. Tyner[1], Matúš Adamkovič[6,7,8], Shilaan Alzahawi[9], Saule Anafinova[10,11], Eli Awtrey[12], Erick Axxe[13], James Bailey[14], Bert N. Bakker[15], Akshaya Balaji[16], Gabriel Banik[17], František Bartoš[15], Henk Berkman[18], Zachariah Berry[19], Felix S. Bethke[20], Timothy F. Brady[21], Nate Breznau[22], Sara Capitan[23], Tabaré Capitán[23], Laura Caquelin[24], Kent Jason Cheng[25], William J. Chopik[26], Gwen-Jiro Clochard[27], Tom Coupé[28,29], Jamie Cummins[30], Elif Gizem Demirag Burak[31], Jianhua Duan[28,32], Kevin M. Esterling[33], Thomas R. Evans[34], Nathan Fiala[35], James Field[36], Victor Gay[37], Jing Geng[38], Johanna Gereke[39,40], Ilka Helene Gleibs[41], Amélie Gourdon-Kanhukamwe[42,43], Dmitry Grigoryev[44], Nicholas Gunby[29,45], Paul H. P. Hanel[46], Sanghyun Hong[28], Sean Dae Houlihan[47], Nick Huntington-Klein[48], Kamil Izydorczak[49], Kristin Jankowsky[50], Kai Jonas[51], Pavol Kačmár[52], Hansika Kapoor[16,35], Sebastian Karcher[53], Marta Kołczyńska[54], David Kretschmer[55], Ljiljana Lazarevic[56], Katelin E. Leahy[26], Jessica C. Lee[57,58], Christopher Limnios[14], An-Chiao Liu[59], John Wills Lloyd[60], Ruben Lopez-Nicolas[61], Nigel Mantou Lou[62], Richard E. Lucas[26], Maximilian Maier[63], Daniel J. Mallinson[64], Marcel Martončik[7,65], Michael C. McCall[53], Nikita Mehta[16], Esteban Méndez[66], Johannes Michalak[67], Daniel C. Molden[68], Faisal Mushtaq[69], Claudia Neuendorf[70,71], Austin Lee Nichols[72], Gustav Nilsonne[24,73], Ernest O'Boyle[74], Jeewon Oh[53], Thomas Ostermann[67], Abiola Oyebanjo[75], Radoslaw Panczak[30], Yuri G. Pavlov[71], Zoran Pavlović[56], Noemi Peter[76], Kim Peters[77], Nathaniel D. Porter[38], Mariah Purol[78], Arathy Puthillam[16,79], Marco Ramljak[59,80], Arran T. Reader[81], W. Robert Reed[28,29], Jan Philipp Röer[67], Ivan Ropovik[8,82,83], Alexander O. Savi[15], Kathleen Schmidt[84,85], Landon Schnabel[86], Eric L. Sevigny[87], Samuel Shaki[88], Shishir Shakya[89], Andrew Soh[90], Angela Somo[91], Fatih Sonmez[92], Eirik Strømland[93], Jordan W. Suchow[94], Anna Szabelska[95], Anirudh Tagat[16], Melba Verra Tutor[96], Karolina Urbanska[97], Pieter Van Dessel[98], Elisabeth Julie Vargo[5], Diem Thi Hong Vo[29,99], Victor Volkman[35], Ke Wang[60], Aaron L. Wichman[100], Jamal R. Williams[79], Fabian Winter[101], Ferdinand Wintermantel[80,102], Nan Zhang[39], Ignazio Ziano[102], Cristina Zogmaister[103], Zorana Zupan[56], Brian A. Nosek[1,60] ✉ & Timothy M. Errington[1]

[1]Center for Open Science Charlottesville, Charlottesville, VA, USA. [2]Wageningen University and Research, Wageningen, The Netherlands. [3]Massachusetts Institute of Technology, Cambridge, MA, USA. [4]University of Coimbra, Coimbra, Portugal. [5]Institute for Globally Distributed Open Research and Education, Stockholm, Sweden. [6]Slovak Academy of Sciences, Bratislava, Slovakia. [7]University of Jyväskylä, Jyväskylä, Finland. [8]Charles University, Prague, Czech Republic. [9]Stanford University, Stanford, CA, USA. [10]Budapest University of Technology and Economics, Budapest, Hungary. [11]Kazakh National University of Water Management and Irrigation, Taraz, Kazakhstan. [12]University of Cincinnati, Cincinnati, OH, USA. [13]Hendrix College, Conway, AR, USA. [14]Providence College, Providence, RI, USA. [15]University of Amsterdam, Amsterdam, The Netherlands. [16]Monk Prayogshala, Mumbai, India. [17]Pavol Jozef Safarik University, Kosice, Slovakia. [18]University of Auckland, Auckland, New Zealand. [19]University of Southern California, Los Angeles, CA, USA. [20]Peace Research Institute Frankfurt (PRIF), Hesse, Germany. [21]University of California, San Diego, La Jolla, CA, USA. [22]German Institute for Adult Education—Leibniz Institute for Lifelong Learning, Bonn, Germany. [23]Swedish University of Agricultural Sciences, Uppsala, Sweden. [24]Karolinska Institutet, Stockholm, Sweden. [25]University of Massachusetts Boston, Boston, MA, USA. [26]Michigan State University, East Lansing, MI, USA. [27]Osaka University, Ibaraki, Japan. [28]University of Canterbury, Christchurch, New Zealand. [29]UCMeta, Christchurch, New Zealand. [30]University of Bern, Bern, Switzerland. [31]University of Oklahoma, Norman, OK, USA. [32]Stats NZ, Christchurch, New Zealand. [33]University of California Riverside, Riverside, CA, USA. [34]University of Greenwich, London, UK. [35]University of Connecticut, Storrs, CT, USA. [36]West Virginia University, Morgantown, WV, USA. [37]Toulouse School of Economics, Toulouse, France. [38]Virginia Tech, Blacksburg, VA, USA. [39]University of Mannheim, Mannheim, Germany. [40]Federal Institute for Population Research (BiB), Wiesbaden, Germany. [41]London School of Economics, London, UK. [42]Kingston University London, London, UK. [43]King's College London, London, UK. [44]HSE University, Moscow, Russia. [45]Contact Energy, Wellington, New Zealand. [46]University of Essex, Colchester, UK. [47]Dartmouth College, Hanover, NH, USA. [48]Seattle University, Seattle, WA, USA. [49]SWPS University, Warszawa, Poland. [50]University of Kassel, Kassel, Germany. [51]Maastricht University, Maastricht, The Netherlands. [52]Faculty of Arts, Pavol Jozef Šafárik University in Košice, Košice, Slovakia. [53]Syracuse University, Syracuse, NY, USA. [54]Institute of Political Studies of the Polish Academy of Sciences, Warszawa, Poland. [55]University of Oxford, Oxford, UK. [56]Faculty of Philosophy, University of Belgrade, Belgrade, Serbia. [57]University of Sydney, Sydney, New South Wales, Australia. [58]University of New South Wales, Sydney, New South Wales, Australia. [59]Utrecht University, Utrecht, The Netherlands. [60]University of Virginia, Charlottesville, VA, USA. [61]University of Murcia, Murcia, Spain. [62]University of Victoria, Victoria, British Columbia, Canada. [63]University College London, London, UK. [64]Penn State Harrisburg, Middletown, PA, USA. [65]Institute of Social Sciences CSPS SAS, Košice, Slovakia. [66]Central Bank of Costa Rica, San José, Costa Rica.

[67]Witten/Herdecke University, Witten, Germany. [68]Northwestern University, Evanston, IL, USA. [69]University of Leeds, Leeds, UK. [70]University of Potsdam, Potsdam, Germany. [71]University of Tübingen, Tübingen, Germany. [72]Central European University, Vienna, Austria. [73]Stockholm University, Stockholm, Sweden. [74]Indiana University, Bloomington, IN, USA. [75]Policy Innovation Center, Abuja, Nigeria. [76]University of Groningen, Groningen, The Netherlands. [77]University of Exeter, Exeter, UK. [78]Union College, Schenectady, NY, USA. [79]UC San Diego, La Jolla, CA, USA. [80]Zeppelin Universität, Friedrichshafen, Germany. [81]University of Stirling, Stirling, UK. [82]Czech Academy of Sciences, Prague, Czech Republic. [83]University of Presov, Presov, Slovakia. [84]Southern Illinois University, Carbondale, IL, USA. [85]Ashland University, Ashland, OH, USA. [86]Cornell University, Ithaca, NY, USA. [87]Georgia State University, Atlanta, GA, USA. [88]Ariel University, Ariel, Israel. [89]Appalachian State University, Boone, NC, USA. [90]Ateneo de Manila University, Quezon City, Philippines. [91]San Diego State University, San Diego, CA, USA. [92]Muş Alparslan University, Muş, Turkey. [93]Western Norway University of Applied Sciences, Bergen, Norway. [94]Stevens Institute of Technology, Hoboken, NJ, USA. [95]Psychological Science Accelerator, Ashland, OH, USA. [96]Independent Researcher, Quezon City, Philippines. [97]Independent Researcher, Sheffield, UK. [98]Ghent University, Ghent, Belgium. [99]RMIT University Vietnam, Ho Chi Minh City, Vietnam. [100]Western Kentucky University, Bowling Green, KY, USA. [101]University of Zurich, Zurich, Switzerland. [102]Humboldt-Universität zu Berlin, Berlin, Germany. [103]Università degli Studi di Milano-Bicocca, Milan, Italy. ✉e-mail: nosek@cos.io

Methods

We examined whether author data were available so that a reproduction could be attempted, and reproducibility was assessed on the basis of whether the same outcomes as reported originally were observed after conducting the same analysis on the same data.

This reproduction project was part of the DARPA SCORE program to generate and evaluate automated measures of confidence in research claims[25]. Evidence for reproducibility (same analysis, same data) was gathered as a secondary criterion of credibility for the program. Human and machine methods were evaluated on their assessments of replicability (same question, new data)[27,52]. Data, materials, code and other outputs from the program that can be shared without violating privacy or intellectual property restrictions are organized and publicly accessible for evaluation and re-use. This section summarizes the methods of sampling, conducting the reproducibility assessments, aggregating the data across reproducibility assessments and evaluating reproduction outcomes. Further details of these methods are provided in Supplementary Tables 1 and 2.

Sampling and selection of claims

Research claims were identified using a systematic selection process to reduce selection effects and to enhance generalizability to quantitative social and behavioural research. The project started with a sample of 3,000 papers selected by stratified random sampling of a larger set of papers to ensure representativeness across the 62 journals and publication dates from 2009 to 2018. The time period was defined as the 10 years before the project onset, and the journals were selected through an informal review and nomination process among authors of this paper and other researchers. We selected journals that were well regarded, published quantitative research, published a sufficient volume of papers during the time period, and collectively represented the diversity of fields and quantitative approaches in the social and behavioural sciences.

Within each selected journal, we aimed to extract a single claim from each of five papers per year across the 10-year sampling frame, producing approximately 50 claims per journal depending on the availability of eligible papers. Each paper was reviewed by a trained coder who assessed whether the paper was eligible for SCORE. Eligible papers reported at least one inferential test using human or social data and reported a statistically significant test result that supported a claim made in the paper's abstract[52]. Journals with papers that did not produce an eligible claim in the given year were resampled from the same journal and year until five claims were extracted, or there were no more eligible papers. This process yielded 3,000 claims from 3,000 papers across the sample.

From the pool of 3,000 papers, 600 were randomly selected as the papers eligible for conducting reproduction attempts with a similar stratified random sampling process to maintain representativeness. Within this pool of 600 papers, 200 were non-randomly sampled for additional coding. In this subset, we extracted all of the main claims regardless of evidence type (that is, including non-inferential and non-significant evidence). These papers were selected because it appeared likely that we could attempt replications and reproductions of their findings based on feasibility and likely availability of relevant materials, with some adjustments made for representativeness (see ref. 27 for details on the sampling and selection process). Other papers and data were gathered during the SCORE program, but they did not include reproduction attempts and are not discussed in this paper.

Data and code availability

We assessed data and code availability for all 600 papers in our stratified random sample. We coded contextual information about the search for data and code sharing, such as where it was found, whether it was linked to or referenced from within the article, and whether the paper stated that the data were restricted. Coders first carried out a brief review of the paper looking for links or references to supplemental materials that included data or code. If either data or code were not located from the paper, coders searched for publicly available materials online, checking online sources such as the website of the publisher or journal where the paper was published, common online repositories, and personal or laboratory websites of authors. If either data or code were not found, then we emailed the corresponding author and requested the missing content. Retrieved or shared data and code were added to a private OSF project for that paper in preparation for reproducibility assessment.

Reproducibility assessment

In total, 191 papers were eligible for reproduction attempts because we had both author data and code, only author data or source data that could be reconstructed to recreate author data. Of these, 143 papers were assessed for reproducibility. Here, random sampling is lost because selection for reproducibility assessment depends on data availability.

Papers were made available for analyst collaborators to conduct a reproduction attempt. Analysts agreed to attempt reproductions on the basis of factors such as familiarity with the methods, analytical software and topical area. Reproduction teams preregistered the inference criteria for judging success. Reproductions conducted during the first half of the program, without the author code, also preregistered their analysis plans. These plans were put through a peer-review process managed by an independent editor; otherwise, the preregistration documents were reviewed internally by the project coordinators. Approved preregistrations were registered on the OSF before the reproduction attempts were conducted. For reproductions conducted during the second half of the program, we eliminated the preregistration and review of analysis plans and added a transparency report of the reproduction process.

Completed reproduction reports went through an internal quality control review. Data, materials and code were archived on the OSF and made openly available to the maximum extent allowed without violating the privacy of the participants or intellectual property licences for any original content.

Data aggregation

Occasionally ($n = 61$ claims from 49 papers), more than one analyst team conducted a reproduction of the same claim. For reporting purposes, we filtered multiple reproductions through a sequence of decision rules to arrive at a singular outcome for reproducibility. The decision rules were maximally generous to achieving reproducibility. First, we selected whichever reproduction attempt produced outcomes closest to the original, using the reproducibility thresholds of precisely, approximately and not reproduced ($n = 21$ claims). Second, if multiple attempts produced equally close results, we selected the attempt that relied most heavily on the authors' materials ($n = 11$ claims). Third, if multiple attempts produced equally close results with the same materials, then we selected the attempt that was part of a reproduction of multiple claims in the same paper ($n = 23$ claims). Finally, if there were multiple reproductions meeting the prior criterion, then we selected randomly among them ($n = 6$ claims).

Data analysis and inference

Presented statistics are mostly descriptive statistics and precision estimates, using two-tailed 95% CIs. Code used to generate each statistic reported in this paper is provided in the data and code repository (https://doi.org/10.17605/osf.io/ed8pj). Unless otherwise specified, confidence intervals for proportion claims-level analyses are two-tailed CI estimates using the Wilson interval method.

For paper-level analyses, multiple reproduced claims in a single paper are weighted to the paper level to adjust the target population, and clustered at the paper level to account for interclass correlation. If there are 3 reproduction attempts for a given paper, for example, each is weighted to 1/3. Unless otherwise indicated, confidence intervals are estimated through a simple clustered bootstrap clustered at the paper level, whereby 95% CI are estimated for the 2.5 and 97.5 percentile intervals of the bootstrapped sample distribution, using 1,000 bootstrap iterations for every statistic.

For data availability, only one claim per paper was assessed, so no weighting or clustering was necessary. For reproducibility, weighting was pegged to the full population of claims and papers that were assessed for reproducibility including for subset analyses by field and year. Unless specified otherwise, the analyses were weighted to represent the complete collection of claims gathered from the 600 papers included in our randomized sample.

Inclusion and ethics

Researchers from more than 24 nations participated in conducting reproductions. Joining the collaboration was an open process, promoted through social media, primarily by the Center for Open Science and the corresponding author. A variety of roles were defined to maximize opportunities for researchers with varying skills, areas of interest and access to resources to participate. Criteria for earning co-authorship were defined in advance so that researchers could make informed decisions about joining the collaboration. All reproduction studies reported in this Article involved secondary analysis of data of organizations, firms or human participants. None involved primary data collection from human participants; all of the reproduction studies were considered non-human-subject research by ethics review boards (BRANY SBER IRB protocol 20-030-749, protocol 20-019-749 and protocol 21-056-749; concurrence from MRDC HRPO and NIWC-PAC HRPO).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data, materials and code associated with this research that can be shared without restriction are publicly available in a living OSF repository (https://doi.org/10.17605/osf.io/ed8pj). The living OSF repository represents improvements, fixes and additions that occur post-publication. Readers can also access a registered, archived version of this repository that is precisely the data, code and documentation as they existed on publication of this paper (https://doi.org/10.17605/osf.io/kmvst). The repository includes all available documentation for reproduction attempts, regardless of whether they were completed. This includes most of the data and code from the individual reproduction attempts, save for any data that are proprietary or protected that will not be made available, or for which analyst teams were uncertain or unable to confirm that they were allowed to share secondary data. It is possible that some data, materials or code that could be shared openly is not available at the time of publication. Readers are encouraged to contact the corresponding author or the authors of the relevant subproject (Supplementary Table 2) to see if more research content can be shared in the living repository. This paper is part of a collection of papers reporting on the SCORE program. Documentation, data and code for the entire program is available at the OSF (https://doi.org/10.17605/osf.io/dtzx4).

Code availability

Code for individual reproduction projects is available alongside data and materials for each project in the OSF repository (https://doi.org/10.17605/osf.io/ed8pj). This includes a push button package with all code and data used to produce all statistics, figures and tables, and code that populates them directly into the manuscript from a template. A registered, archived version of the repository containing precisely the data, code and documentation used to generate the outcomes reported in this paper is also available at OSF (https://doi.org/10.17605/osf.io/kmvst).

52. Tyner, A. H. et al. Investigating the replicability of the social and behavioural sciences. *Nature* https://doi.org/10.1038/s41586-025-10078-y (2026).

Acknowledgements We thank B. Arendt, A. Denis, S. Field, Z. Loomas, B. Luis, L. Markham, E. S. Parsons, C. Soderberg and A. Russell for their contributions to this project. This work was supported by the Defense Advanced Research Projects Agency (DARPA) under cooperative agreement numbers N660011924015 (principal investigator, B.A.N.) and HR00112020015 (principal investigator, T.M.E.). The views, opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

Author contributions Conceptualization: M.K.S., A.H.T., W.J.C., E.G.D.B., K.E.L., B.A.N. and T.M.E. Data curation: A.L.A., A.H.T., S. Anafinova, K.M.E., D.G. and J.W.L. Formal analysis: M.K.S., A.H.T., M.A., S. Alzahawi, S. Anafinova, E. Axxe, A.B., F.B., Z.B., F.S.B., T.F.B., T. Capitán, L.C., K.J.C., W.J.C., G.-J.C., T. Coupé, J.D., T.R.E., N. Fiala, J.F., V.G., J. Gereke, I.H.G., D.G., N.G., P.H.P.H., S.H., S.D.H., K.I., K. Jankowsky, P.K., M.K., D.K., K.E.L., J.C.L., C.L., A.-C.L., R.L.-N., N.M.L., M. Maier, D.J.M., M. Martončik, N.M., E.M., D.C.M., F.M., C.N., J.O., T.O., A.O., R.P., Y.G.P., Z.P., N.P., N.D.P., M.P., M.R., A.T.R., W.R.R., J.P.R., I.R., A.O.S., K.S., L.S., E.L.S., S. Shaki, A. Somo, F.S., A. Szabelska, A.T., K.U., P.V.D., D.T.H.V., V.V., K.W., A.L.W., J.R.W., F. Wintermantel and N.Z. Funding acquisition: B.A.N. and T.M.E. Investigation: O.M., A.L.A., M. Daley, N. Fox, K.M.H., M.K.S., B.M., A.H.T., M.A., S. Alzahawi, E. Axxe, J.B., G.B., F.B., Z.B., T.F.B., N.B., S.C., T. Capitán, K.J.C., W.J.C., T. Coupé, J.C., E.G.D.B., J.D., T.R.E., N. Fiala, J.F., V.G., J. Gereke, I.H.G., P.H.P.H., S.H., S.D.H., N.H.-K., K. Jankowsky, M.K., C.L., A.-C.L., R.L.-N., D.J.M., M. Martončik, J.M., D.C.M., E.O., J.O., T.O., Y.G.P., Z.P., N.P., N.D.P., M.P., A.P., A.T.R., W.R.R., J.P.R., I.R., A.O.S., K.S., L.S., E.L.S., A. Soh, A. Somo, J.W.S., A. Szabelska, A.T., M.V.T., K.U., P.V.D., D.T.H.V., V.V., K.W., A.L.W., J.R.W., F. Winter, N.Z. and T.M.E. Methodology: O.M., A.L.A., N. Fox, M.K.S., B.M., P.S., A.H.T., M.A., S. Alzahawi, E. Axxe, J.B., F.B., Z.B., T.F.B., N.B., S.C., T. Capitán, W.J.C., T. Coupé, K.M.E., T.R.E., N. Fiala, J.F., V.G., J. Gereke, I.H.G., P.H.P.H., S.D.H., K. Jankowsky, K.E.L., J.C.L., C.L., A.-C.L., R.L.-N., M. Maier, D.J.M., D.C.M., C.N., A.L.N., J.O., Y.G.P., Z.P., N.D.P., M.P., M.R., A.T.R., J.P.R., I.R., A.O.S., K.S., L.S., E.L.S., S. Shaki, S. Shakya, A. Somo, F.S., J.W.S., A. Szabelska, A.T., D.T.H.V., K.W., A.L.W., J.R.W., B.A.N. and T.M.E. Project administration: O.M., A.H.T., B.A.N. and T.M.E. Software: A.L.A., M.K.S., T.S., A.H.T., S. Alzahawi, E. Axxe, J.B., A.B., F.B., T.F.B., N.B., T. Coupé, T.R.E., J.F., V.G., J. Geng, I.H.G., D.G., P.H.P.H., S.D.H., K. Jankowsky, P.K., D.K., K.E.L., C.L., A.-C.L., R.L.-N., D.J.M., N.M., D.C.M., F.M., T.O., A.O., Y.G.P., Z.P., N.D.P., M.P., A.P., M.R., J.P.R., I.R., A.O.S., L.S., E.L.S., S. Shakya, A. Somo, F.S., J.W.S., A.T., M.V.T., D.T.H.V., V.V., K.W., A.L.W., F. Winter and F. Wintermantel. Supervision: O.M., K.M.H., M.K.S., A.H.T., B.N.B., T. Capitán, W.J.C., E.G.D.B., K.M.E., N. Fiala, I.H.G., A.G.-K., K. Jonas, J.W.L., R.E.L., G.N., J.O., N.D.P., W.R.R., J.P.R., S. Shaki, F. Wintermantel, C.Z., B.A.N. and T.M.E. Validation: O.M., A.L.A., N.H., P.S., A.H.T., S. Anafinova, E. Awtrey, E. Axxe, J.B., B.N.B., H.B., N.B., T. Capitán, L.C., W.J.C., J.C., E.G.D.B., K.M.E., I.H.G., A.G.-K., D.G., N.H.-K., K. Jonas, H.K., S.K., K.E.L., C.L., J.W.L., R.E.L., M. Maier, M.C.M., C.N., A.L.N., G.N., J.O., T.O., N.P., K.P., N.D.P., A.T.R., J.P.R., L.S., S. Shaki, E.S., A. Szabelska, A.T., K.U., E.J.V., V.V., F. Wintermantel, I.Z., C.Z., Z.Z. and T.M.E. Visualization: N.H., B.M., A.H.T., N.B., M.R., S. Shaki and B.A.N. Original draft: O.M., B.M., A.H.T., S. Anafinova, W.J.C., N. Fiala, K.E.L., J.O., M.P., S. Shaki, V.V., B.A.N. and T.M.E. Review and editing: O.M., M. Daley, M. Dirzo, B.M., P.S., A.H.T., S. Anafinova, B.N.B., G.B., N.B., W.J.C., J.C., E.G.D.B., N. Fiala, J.F., D.G., P.H.P.H., K. Jankowsky, P.K., S.K., L.L., K.E.L., R.L.-N., N.M.L., M. Maier, M. Martončik, C.N., T.O., R.P., Y.G.P., Z.P., N.P., M.P., A.T.R., J.P.R., I.R., L.S., E.L.S., S. Shaki, S. Shakya, E.S., J.W.S., V.V., K.W., A.L.W., B.A.N. and T.M.E.

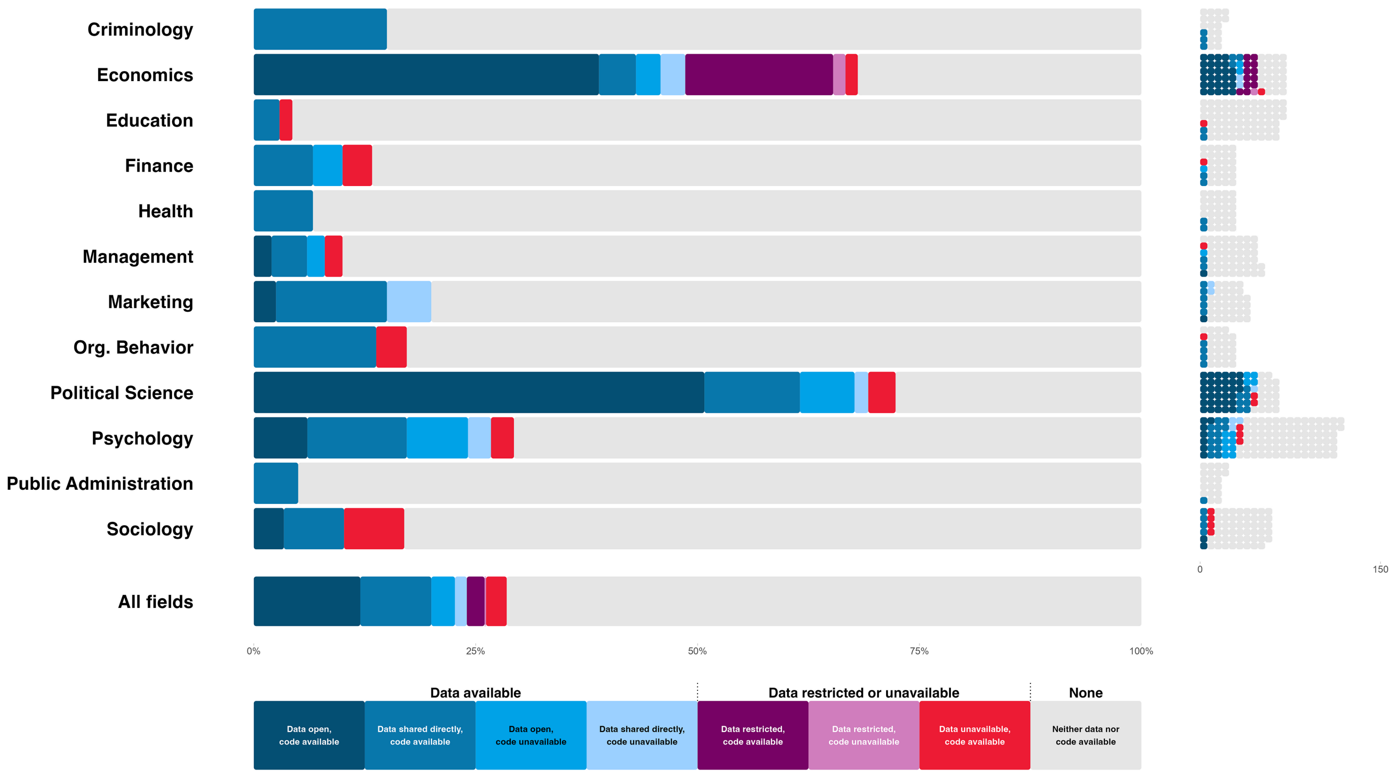
Competing interests A.H.T., M. Daley, N.H., K.M.H., O.M., T.S., B.A.N. and T.M.E. are employees of the non-profit organization Center for Open Science that has a mission to increase openness, integrity and trustworthiness of research.

Additional information

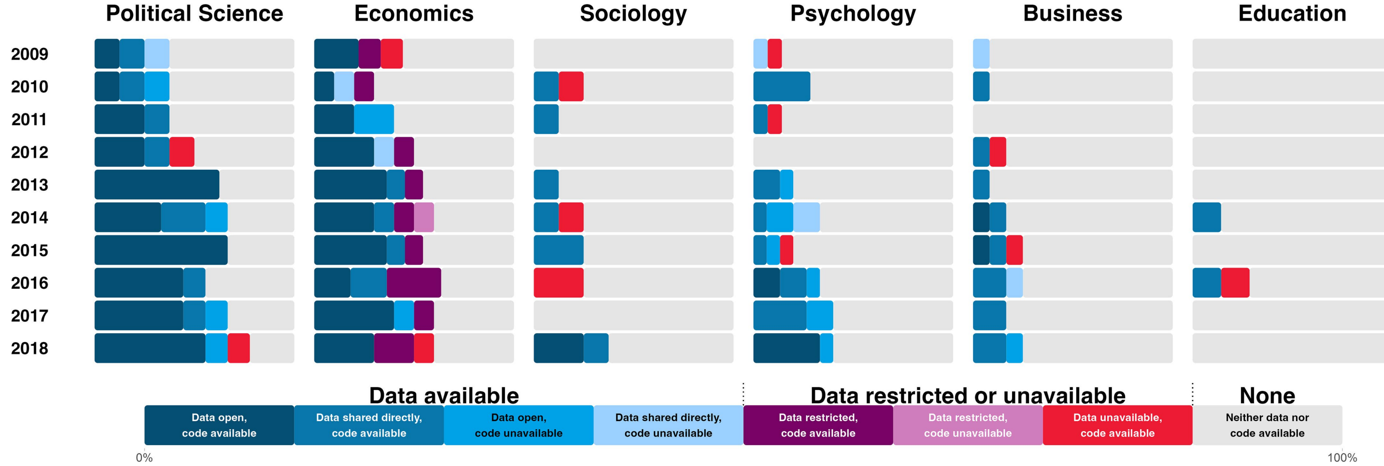
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-10203-5.

Correspondence and requests for materials should be addressed to Brian A. Nosek.

Peer review information *Nature* thanks Alfredo Sánchez-Tójar and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

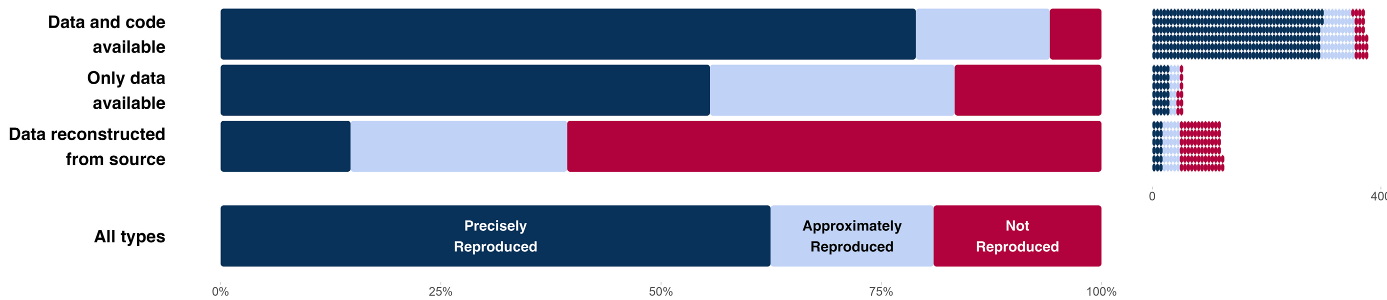


Extended Data Fig. 1 | Data availability rates by 12 subfields. The left panel shows data and code availability as a percentage of papers; the right panel shows raw counts of papers with data and code available and not available. Note that purple reflects restricted data, which did not count as available data, but might be accessible in principle. This is presented as Fig. S6 in the Supporting Information with additional narrative context.

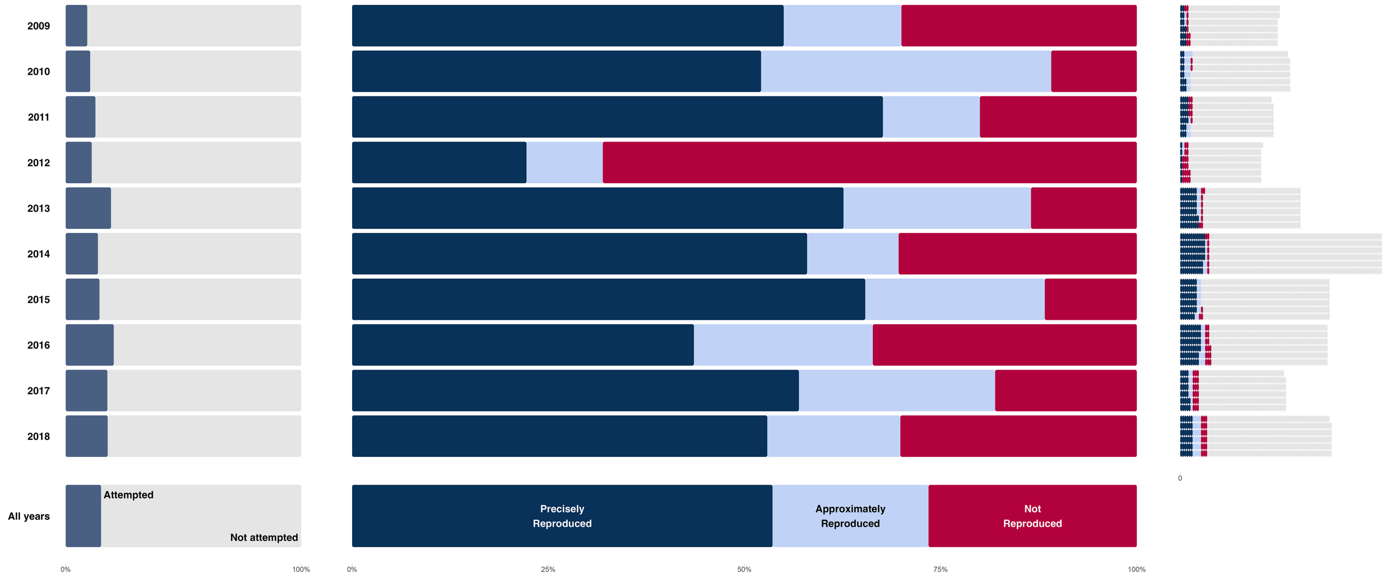


Extended Data Fig. 2 | Data availability rates by year of publication for all fields. Smallest sample sizes per cell were in Education (n's from 6 to 7 per year). Note that purple reflects restricted data, which did not count as available data, but might be accessible in principle. This is presented as Fig. S7 in the Supporting Information with additional narrative context.

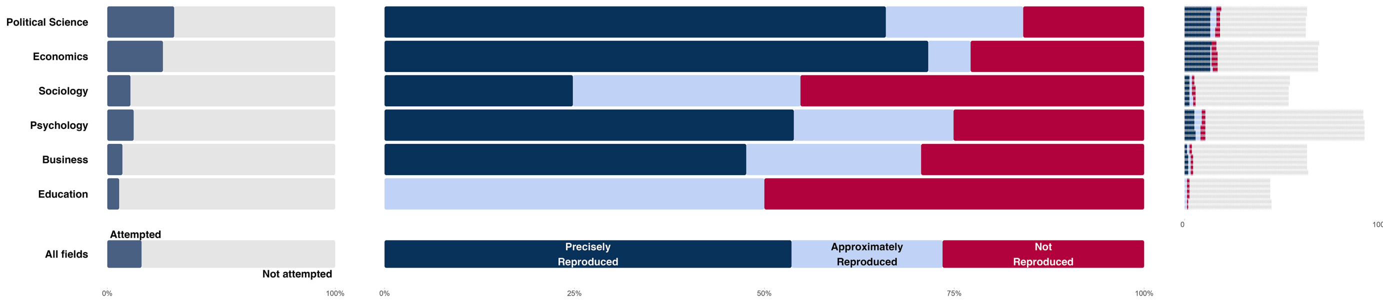
Extended Data Fig. 3 | Data availability rates by year of publication for 12 subfields. Smallest sample sizes per cell were in criminology and public administration, each having an n of 2 per year. Note that purple reflects restricted data, which did not count as available data, but might be accessible in principle. This is presented as Fig. S8 in the Supporting Information with additional narrative context.



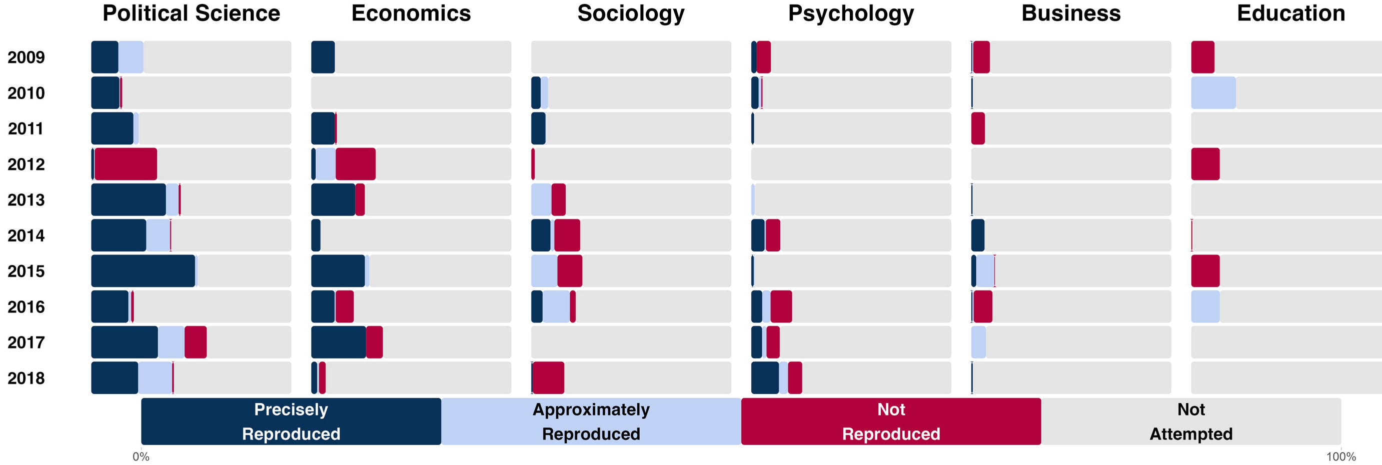
Extended Data Fig. 4 | Reproducibility by whether data and code were available, only data were available, or when the paper's data were reconstructed from available source data for all claims. Reproducibility success rates as a percentage of attempts (left), and reproducibility success rates as counts (right). This is presented as Fig. S1 in the Supporting Information with additional narrative context.



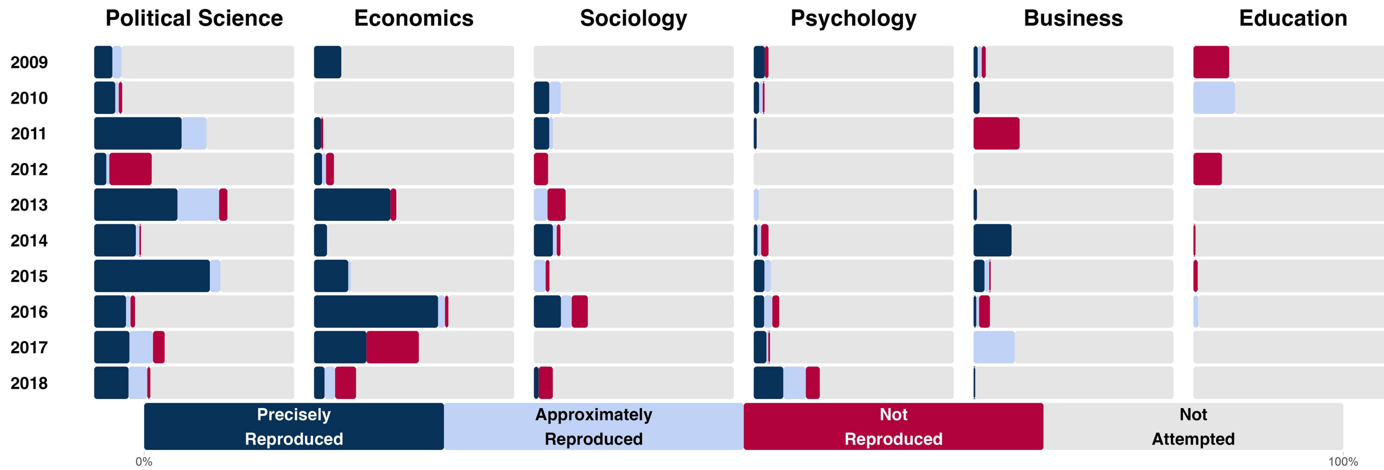
Extended Data Fig. 5 | Reproducibility by year of publication for all claims. The left column illustrates the proportion of outcome reproduction attempts from the sample of claims. The middle column illustrates reproducibility as a percentage of the attempts. The right column illustrates reproducibility as counts compared with the sample of claims. This is presented as Fig. S2 in the Supporting Information with additional narrative context.



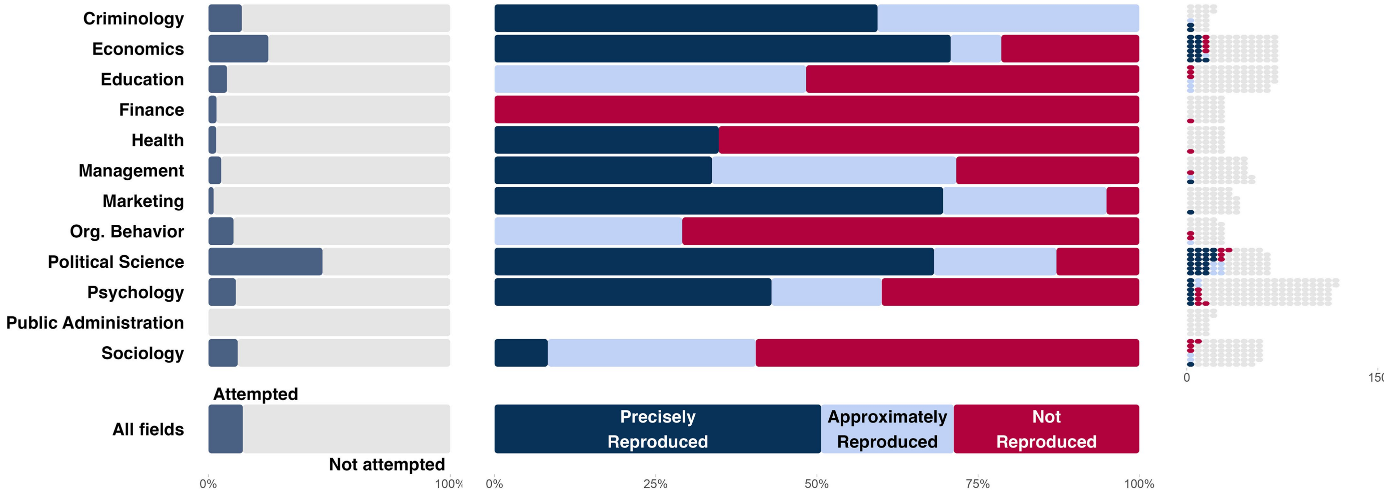
Extended Data Fig. 6 | Reproducibility by field for all claims. The left column illustrates the proportion of outcome reproduction attempts from the sample of claims. The middle column illustrates reproducibility as a percentage of the attempts. The right column illustrates reproducibility as counts compared with the sample of claims. This is presented as Fig. S3 in the Supporting Information with additional narrative context.



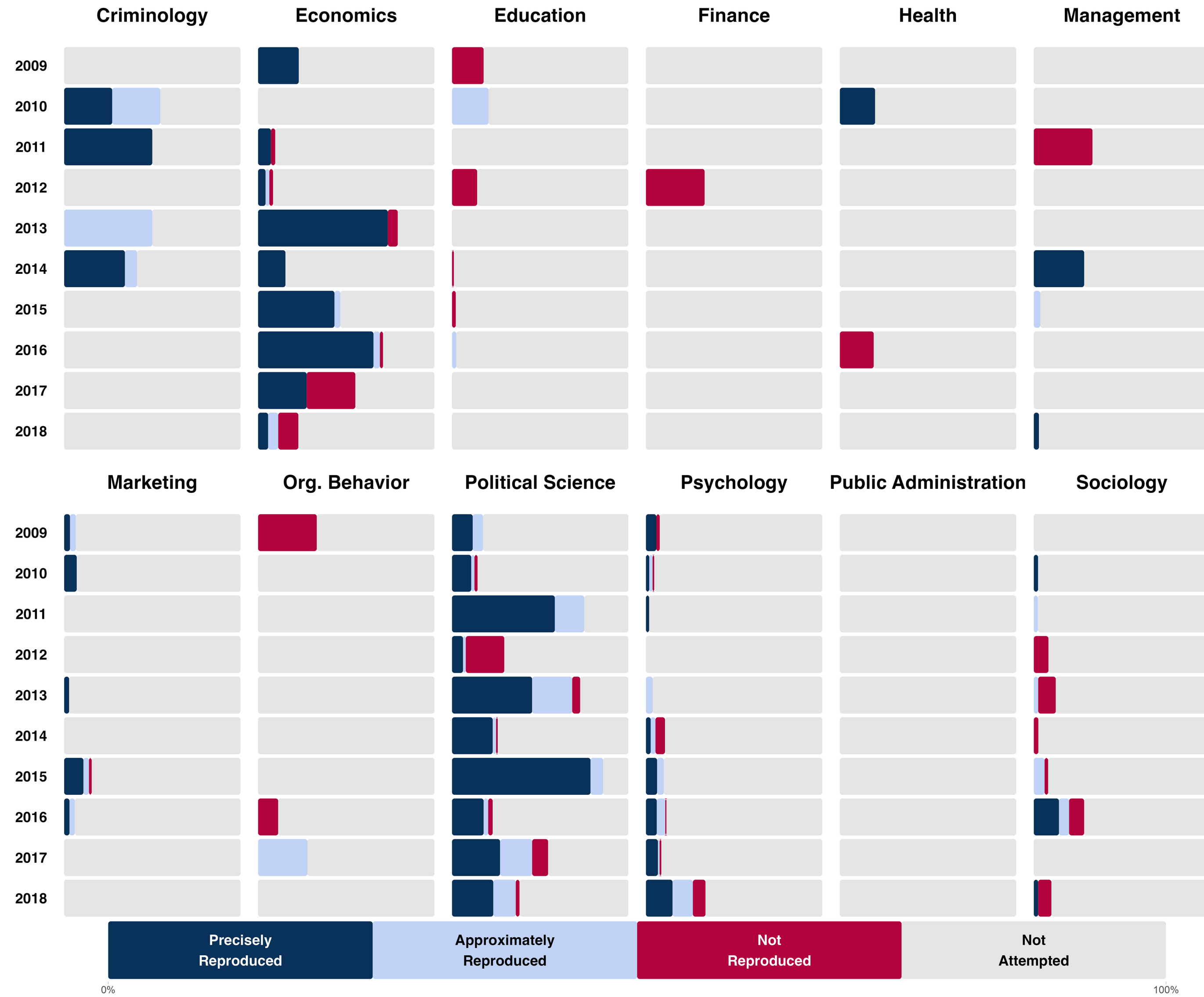
Extended Data Fig. 7 | Reproducibility by field and year by paper as a proportion of the sample. Reproducibility as a percentage of the sample of papers from each year and each field. This is presented as Fig. S4 in the Supporting Information with additional narrative context.



Extended Data Fig. 8 | Reproducibility by field and year by claim as a proportion of the sample. Reproducibility as a percentage of the sample of claims from each year and each field. This is presented as Fig. S5 in the Supporting Information with additional narrative context.



Extended Data Fig. 9 | Reproducibility by 12 subfields. The left column illustrates the proportion of outcome reproduction attempts from the sample of papers. The middle column illustrates reproducibility as a percentage of the attempts. The right column illustrates reproducibility as counts compared with the sample of papers. This is presented as Fig. S9 in the Supporting Information with additional narrative context.



Extended Data Fig. 10 | Reproducibility by 12 subfields and by year. Reproducibility as a percentage of the sample of papers from each year and each field. This is presented as Fig. S10 in the Supporting Information with additional narrative context.

nature portfolio

Corresponding author(s): Brian Nosek

Last updated by author(s): Jan 20, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: N/A

Data analysis: All analyses of the collected reproduction outcomes were performed in R (4.5.0 (2024-10-31)). The full list of loaded packages (including dependencies) is as follows:

- abind [* -> 1.4-8]
- askpass [* -> 1.2.1]
- assertthat [* -> 0.2.1]
- backports [* -> 1.5.0]
- base64enc [* -> 0.1-3]
- bayesplot [* -> 1.14.0]
- BH [* -> 1.87.0-1]
- bit [* -> 4.6.0]
- bit64 [* -> 4.6.0-1]
- blob [* -> 1.2.4]
- boot [* -> 1.3-32]
- bridgesampling [* -> 1.1-2]
- brms [* -> 2.23.0]
- Brobdingnag [* -> 1.2-9]
- broom [* -> 1.0.10]
- broom.mixed [* -> 0.2.9.6]

```
- bslib          [* -> 0.9.0]
- cachem         [* -> 1.1.0]
- callr          [* -> 3.7.6]
- cellranger     [* -> 1.1.0]
- checkmate      [* -> 2.3.3]
- class          [* -> 7.3-23]
- cli            [* -> 3.6.5]
- clipr          [* -> 0.8.0]
- cluster        [* -> 2.1.8.1]
- coda           [* -> 0.19-4.1]
- codetools      [* -> 0.2-20]
- colorspace     [* -> 2.1-2]
- colourpicker   [* -> 1.3.0]
- commonmark     [* -> 2.0.0]
- conflicted     [* -> 1.2.0]
- corrplot       [* -> 0.95]
- cowplot        [* -> 1.2.0]
- cpp11          [* -> 0.5.2]
- crayon         [* -> 1.5.3]
- crosstalk      [* -> 1.2.2]
- curl           [* -> 7.0.0]
- data.table     [* -> 1.17.8]
- DBI            [* -> 1.2.3]
- dbplyr         [* -> 2.5.1]
- desc           [* -> 1.4.3]
- DescTools      [* -> 0.99.60]
- digest         [* -> 0.6.37]
- distributional [* -> 0.5.0]
- dplyr          [* -> 1.1.4]
- DT             [* -> 0.34.0]
- dtplyr         [* -> 1.3.2]
- e1071          [* -> 1.7-16]
- evaluate       [* -> 1.0.5]
- evmix          [* -> 2.12]
- Exact          [* -> 3.3]
- expm           [* -> 1.0-0]
- farver         [* -> 2.1.2]
- fastmap        [* -> 1.2.0]
- fontawesome    [* -> 0.5.3]
- forcats        [* -> 1.0.1]
- foreach        [* -> 1.5.2]
- foreign        [* -> 0.8-90]
- Formula        [* -> 1.2-5]
- fs             [* -> 1.6.6]
- funkyheatmap   [* -> 0.5.2]
- furrr          [* -> 0.3.1]
- future         [* -> 1.67.0]
- future.apply   [* -> 1.20.0]
- gargle         [* -> 1.6.0]
- gdata          [* -> 3.0.1]
- generics       [* -> 0.1.4]
- ggExtra        [* -> 0.11.0]
- ggforce        [* -> 0.5.0]
- ggplot2        [* -> 4.0.0]
- ggridges       [* -> 0.5.7]
- ggside         [* -> 0.4.0]
- gld            [* -> 2.6.8]
- glmnet         [* -> 4.1-10]
- globals        [* -> 0.18.0]
- glue           [* -> 1.8.0]
- googledrive    [* -> 2.1.2]
- googlesheets4  [* -> 1.1.2]
- gridExtra      [* -> 2.3]
- gsl            [* -> 2.1-8]
- gtable         [* -> 0.3.6]
- gtools         [* -> 3.9.5]
- haven          [* -> 2.5.5]
- highr          [* -> 0.11]
- Hmisc          [* -> 5.2-4]
- hms            [* -> 1.1.4]
- htmlTable      [* -> 2.4.3]
- htmltools      [* -> 0.5.8.1]
- htmlwidgets    [* -> 1.6.4]
- httpuv         [* -> 1.6.16]
- httr           [* -> 1.4.7]
- ids            [* -> 1.0.1]
```

- inline [* -> 0.3.21]
- isoband [* -> 0.2.7]
- iterators [* -> 1.0.14]
- jomo [* -> 2.7-6]
- jquerylib [* -> 0.1.4]
- jsonlite [* -> 2.0.0]
- knitr [* -> 1.50]
- labeling [* -> 0.4.3]
- later [* -> 1.4.4]
- lattice [* -> 0.22-7]
- lazyeval [* -> 0.2.2]
- lifecycle [* -> 1.0.4]
- listenv [* -> 0.9.1]
- lme4 [* -> 1.1-37]
- lmom [* -> 3.2]
- loo [* -> 2.8.0]
- lubridate [* -> 1.9.4]
- magrittr [* -> 2.0.4]
- MASS [* -> 7.3-65]
- Matrix [* -> 1.7-4]
- matrixStats [* -> 1.5.0]
- memoise [* -> 2.0.1]
- mgcv [* -> 1.9-3]
- mice [* -> 3.18.0]
- mime [* -> 0.13]
- miniUI [* -> 0.1.2]
- minqa [* -> 1.2.8]
- mitml [* -> 0.4-5]
- mnormt [* -> 2.1.1]
- modelr [* -> 0.1.11]
- mvtnorm [* -> 1.3-3]
- nleqslv [* -> 3.3.5]
- nlme [* -> 3.1-168]
- nloptr [* -> 2.2.1]
- nnet [* -> 7.3-20]
- numDeriv [* -> 2016.8-1.1]
- officer [* -> 0.7.0]
- openssl [* -> 2.3.4]
- ordinal [* -> 2023.12-4.1]
- otel [* -> 0.2.0]
- pan [* -> 1.9]
- pandoc [* -> 0.2.0]
- parallelly [* -> 1.45.1]
- patchwork [* -> 1.3.2]
- pbapply [* -> 1.7-4]
- pillar [* -> 1.11.1]
- pkgbuild [* -> 1.4.8]
- pkgconfig [* -> 2.0.3]
- plyr [* -> 1.8.9]
- polyclip [* -> 1.10-7]
- posterior [* -> 1.6.1]
- prettyunits [* -> 1.2.0]
- processx [* -> 3.8.6]
- progress [* -> 1.2.3]
- promises [* -> 1.4.0]
- proxy [* -> 0.4-27]
- ps [* -> 1.9.1]
- purrr [* -> 1.1.0]
- QuickJSR [* -> 1.8.1]
- R6 [* -> 2.6.1]
- ragg [* -> 1.5.0]
- rappdirs [* -> 0.3.3]
- rbibutils [* -> 2.3]
- RColorBrewer [* -> 1.1-3]
- Rcpp [* -> 1.1.0]
- RcppArmadillo [* -> 15.0.2-2]
- RcppEigen [* -> 0.3.4.0.2]
- RcppParallel [* -> 5.1.11-1]
- Rdpack [* -> 2.6.4]
- readr [* -> 2.1.5]
- readxl [* -> 1.4.5]
- reformulas [* -> 0.4.1]
- rematch [* -> 2.0.0]
- rematch2 [* -> 2.1.2]
- renv [* -> 1.1.5]
- reprex [* -> 2.1.1]

```
- reshape2      [* -> 1.4.4]
- rlang         [* -> 1.1.6]
- rmarkdown     [* -> 2.30]
- rootSolve     [* -> 1.8.2.4]
- rpart         [* -> 4.1.24]
- rstan         [* -> 2.32.7]
- rstantools    [* -> 2.5.0]
- rstudioapi    [* -> 0.17.1]
- rvest         [* -> 1.0.5]
- S7            [* -> 0.2.0]
- sass          [* -> 0.4.10]
- scales        [* -> 1.4.0]
- selectr       [* -> 0.4-2]
- shape         [* -> 1.4.6.1]
- shiny         [* -> 1.11.1]
- shinyjs       [* -> 2.1.0]
- sourcetools   [* -> 0.1.7-1]
- SparseM       [* -> 1.84-2]
- StanHeaders   [* -> 2.32.10]
- stringi       [* -> 1.8.7]
- stringr       [* -> 1.5.2]
- survival      [* -> 3.8-3]
- sys           [* -> 3.4.3]
- systemfonts   [* -> 1.3.1]
- tensorA       [* -> 0.36.2.1]
- textshaping   [* -> 1.0.4]
- tibble        [* -> 3.3.0]
- tidyr         [* -> 1.3.1]
- tidyselect    [* -> 1.2.1]
- tidyverse     [* -> 2.0.0]
- timechange    [* -> 0.3.0]
- tinytex       [* -> 0.57]
- tweenr        [* -> 2.0.3]
- tzdb          [* -> 0.5.0]
- ucminf        [* -> 1.2.2]
- utf8          [* -> 1.2.6]
- uuid          [* -> 1.2-1]
- vctrs         [* -> 0.6.5]
- viridisLite   [* -> 0.4.2]
- vroom         [* -> 1.6.6]
- wCorr         [* -> 1.9.8]
- weights       [* -> 1.1.2]
- withr         [* -> 3.0.2]
- xfun          [* -> 0.53]
- xml2          [* -> 1.4.1]
- xtable        [* -> 1.8-4]
- yaml          [* -> 2.3.10]
- zcurve        [* -> 2.4.5]
- zip           [* -> 2.3.3]
```

All code will be made available on the Open Science Framework at https://osf.io/ed8pj/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data, materials, and code associated with this research that can be shared without restriction are publicly available in a living OSF repository (https://osf.io/ed8pj/; DOI: 10.17605/OSF.IO/ED8PJ). The living OSF repository represents improvements, fixes, and additions that occur post-publication. Readers can also access a registered, archived version of this repository that is precisely the data, code, and documentation as they existed upon publication of this paper (https://osf.io/4m8xa/; DOI: 10.17605/OSF.IO/4M8XA).

The repository includes all available documentation for reproduction attempts regardless of whether they were completed. This includes most of the data and code from the individual reproduction attempts, save for any data that is proprietary or protected that will not be made available, or for which analyst teams were uncertain or unable to confirm that they were allowed to share secondary data. It is possible that some data, materials, or code that could be shared openly is not available at the time of publication. Readers are encouraged to contact the corresponding author or the authors of the relevant subproject (Table S2) to see if more research content can be shared in the living repository.

This paper is part of a collection of papers reporting on the SCORE program. Documentation, data, and code for the entire program is available at https://osf.io/dtzx4/, DOI: 10.17605/OSF.IO/DTZX4).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a quantitative study reporting data on 164 reproduction studies and availability of reproduction resources for 600 published articles.
Research sample	The sample for this study is aggregated data from reproduction attempts of 557 claims from 148 original articles that were published between 2009-2018 in 62 social and behavioral science journals, alongside assessments of the availability of materials to conduct reproduction attempts from 600 original studies. There is no relevant demographic information to report as this study's units of observations are the individual reproduction outcomes and the original articles. Original studies were made eligible for reproduction through stratified sampling based on the journal and the year of publication, in an attempt to collect a representative sample of articles published in these journals in this 10-year period. Original results were selected for reproduction by the researchers conducting the reproduction attempts, so the resulting sample is based on factors like the match between the original studies and the analysts' research expertise, as well as data availability, timing, and other contingencies. We conducted these reproduction studies to provide a sample of recent reproduction outcomes that cut across the core domains of the social and behavioral sciences.
Sampling strategy	We targeted stratified random selection of 30,000 papers from 62 journals across the social and behavioral sciences to be the largest sample of the published literature that could feasibly be managed with the resources of the program. We randomly selected 3,900 papers from this sample as the largest feasible sample for coding claims that could be evaluated by human and machine assessment teams. We selected 600 papers from 3,000 of them as eligible for reproduction attempts with the recognition that costs of reproduction studies in both time and money would limit the number of reproductions that could be completed, so subsetting the larger sample would increase the feasibility of maintaining representativeness of the reproduction studies with the larger sample.
Data collection	Data collection occurred through a group of independent labs, research groups, and principal investigators who selected the original claims that they wanted to attempt to reproduce. Researchers were not blinded to the original studies' claims/hypotheses that they were attempting to reproduce. Please see the "Sourcing Reproduction Analysts" section of the manuscript for details.
Timing	Analysts collected or prepared data for the individual reproduction projects during the period of active program funding (approximately February 2019 through March 2023). We do not have the precise dates of data collection for the reproduction studies themselves, which were administered by the respective reproduction teams.
Data exclusions	We did not exclude reproduction outcomes from the aggregated data.
Non-participation	N/A
Randomization	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.