SCIENTIFIC COMMUNITY

Nonreplicable publications are cited more than replicable ones

Marta Serra-Garcia*† and Uri Gneezy†

We use publicly available data to show that published papers in top psychology, economics, and general interest journals that fail to replicate are cited more than those that replicate. This difference in citation does not change after the publication of the failure to replicate. Only 12% of postreplication citations of nonreplicable findings acknowledge the replication failure. Existing evidence also shows that experts predict well which papers will be replicated. Given this prediction, why are nonreplicable papers accepted for publication in the first place? A possible answer is that the review team faces a trade-off. When the results are more "interesting," they apply lower standards regarding their reproducibility.

INTRODUCTION

The replication crisis in social sciences refers to the failure to replicate a large fraction of published experiments (*1*) and the selective publication of results and specifications (*2*–*4*). Three influential replication projects (*5*–*7*) tried to systematically replicate the findings in top psychology, economics, and general science journals. In psychology, only 39% of the experiments yielded significant findings in the replication study, compared to 97% of the original experiments. In economics, 61% of 18 studies replicated, and among *Nature/Science* publications, 62% of 21 studies did. In addition, the relative effect sizes of findings that did replicate were only 75% of the original ones. For failed replications, they were close to 0% [see also (*8*–*10*)]. Prediction markets, in which experts in the field bet on the replication results before the replication studies, showed that experts could predict well which findings would replicate (*11*).

Here, we use the findings from these three replication projects to correlate replicability with citations and test whether papers that failed to replicate are cited significantly more often than those that were successfully replicated, both before and after the replication projects were published. We collected two types of measures: (i) replicability measures and prediction market results, which are publicly available for all three replication projects; and (ii) Google Scholar citations from the date of publication until the end of 2019. We additionally collected several proxies for the quality of these citations: how often citations are themselves cited, whether they are published, and the impact factor of the journals in which they are published. We examine the relationship between citations and other measures of impact and replicability across the three replication projects.

The number of citations is a basic measure that is used to assess the scholarly impact of a published work. It is used to study intellectual history and evaluate the quality of scientific work across a variety of disciplines (*12*, *13*). In promotion decisions, for example, most academic institutions use citations as an important metric in the decision of whether to promote a faculty member. Citation is a proxy of the impact of a paper, and with all else being equal, researchers would prefer to be cited more. The proxy is also clearly noisy because papers are cited for a myriad of reasons. In our analysis, we start by examining the correlation between citation counts and replicability. We then examine the relationship between other measures of impact, such as the impact factors of the journals in which citations are published, and replicability. Still, citations could be "negative" in the sense that they mention the original result's failure to replicate. We also explore this possibility by classifying as negative or positive/neutral the type of citation after the replication project is published.

Our main finding is that papers that fail to replicate in (*5*–*7*) are cited more than those that are replicable. We find no significant change in citation trends, even after the publication of the failed replication. Notably, only a minority of publications after the failed replications were published acknowledge the failure. The "quality" of citations of papers that failed to replicate is similar to that of papers that were replicated: We do not find a difference in how often the citations of nonreplicable publications are cited by others or in the impact factor of the journals in which citations are published.

Assuming more cited papers present more "interesting" findings, a negative correlation between replicability and citation count could reflect a review process that is laxer when the results are more interesting (by interesting, we mean papers that attract more attention and follow-up work). Supporting evidence for this explanation is provided in (*11*), which showed that experts in the field successfully predicted which findings would replicate before the replication studies were run. Our analysis shows that experts' predictions regarding which studies will replicate do not merely reflect statistical features (e.g., statistical power) of the original papers. Yet, the publication process involves expert reviewers and editors who allowed these papers to be published, despite the skepticism reflected later in prediction markets (*11*, *14*).

RESULTS

Nonreplicable publications are cited more even after the replication study is published

Figure 1 shows the distribution of total citation counts by the end of 2019 of the papers included in the replication projects, separately depending on replicability for *Nature/Science*, economics journals, and psychology journals. Replicability is measured according to the criterion that the replication project featured a P value of 0.05 or lower in a two-sided test, with an effect in the same direction as the original test.

Rady School of Management, University of California, San Diego La Jolla, CA, USA.
*Corresponding author. Email: mserragarcia@ucsd.edu
†These authors contributed equally to this work.

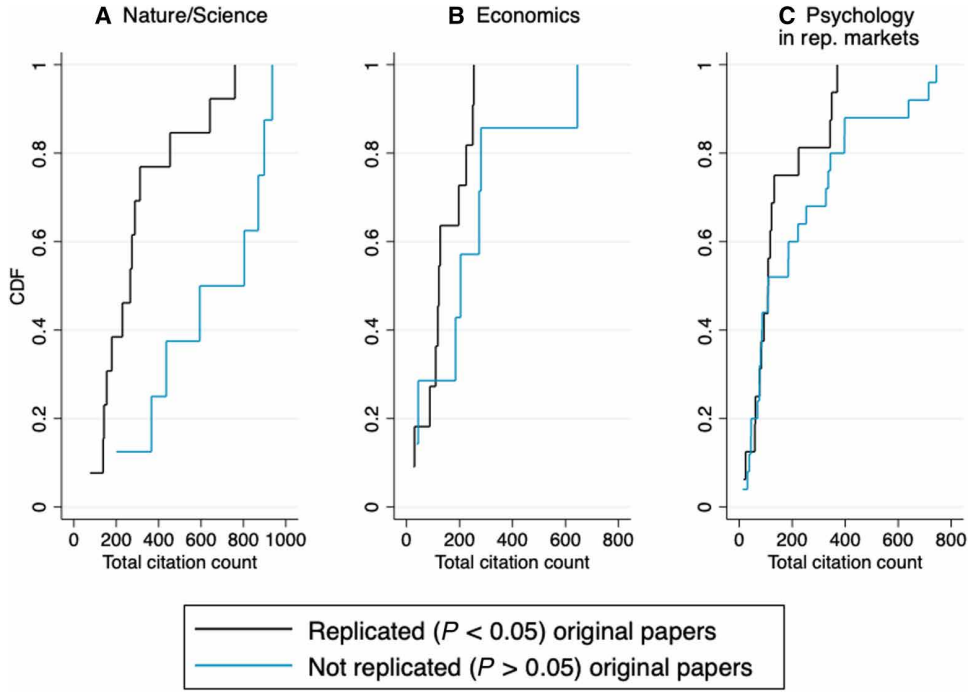

Fig. 1. Distribution of total citation counts and replicability. The distribution of citation counts, separately depending on replicability (a *P* value of 0.05 or lower in a two-sided test and with an effect in the same direction as the original one), is shown separately for *Nature/Science* (**A**), Economics (**B**), and Psychology (**C**), for which replication markets were conducted, as reported in (*11*). CDF, cumulative distribution function.

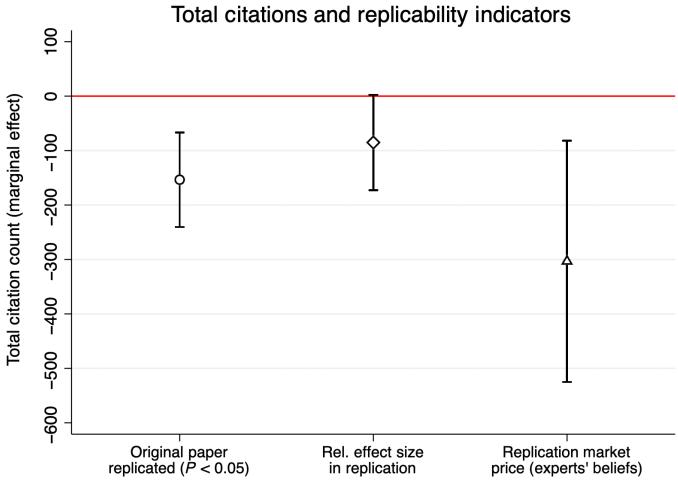

Fig. 2. Total citation count and replicability measures. Average marginal effect on total citation count, as a function of whether the original paper was replicated (leftmost confidence interval), the relative effect size of the replication, and the market price reflecting experts' belief that each study would replicate (between 0 and 1). Effects obtained from Poisson regressions, pooling the three replication projects for which replication markets were conducted, as reported in (*11*), and adding project fixed effects and robust standard errors. The regression table is provided in table S3. Bars indicate 95% confidence intervals.

As shown in Fig. 2, papers that replicate are cited 153 times less, on average, than papers that do not (N = 80, Poisson regression, residual df = 76, Z = −3.47, and P = 0.001). The point estimate for the difference in citations is largest for papers published in *Nature* and *Science*, compared with studies published in economics and psychology journals. Yet, the relationship between replicability and citations is not significantly different across the three replication projects. When we include several individual characteristics of the studies replicated [based on (*15*)], such as the number of authors, the rate of male authors, and the characteristics of the experiment (location, language, and online implementation), as well as the field in which the paper was published (*16*), the relationship between replicability and citations is qualitatively unchanged (the same occurs if we control for the highest seniority level among the authors). In the Supplementary Materials, we provide a robustness analysis based on specification curves that display the robustness of results across 36 different regression models (*17*).

Next, we examine the relationship between two additional measures of replicability and citation counts, as shown in Fig. 2. First, studies with a larger relative effect size of the replication study, compared with the original studies, are cited less, though the difference is only marginally significant (N = 79, Poisson regression, residual df = 75, Z stat = −1.91, and P = 0.057). Moving from a relative effect size of 0 (no replication) to 1 (perfect replication of the original study) is associated with 85 fewer citations, on average.

Second, studies that experts predicted would be less likely to replicate, as reflected by the market price in replication markets, have a higher citation count than those they predicted would replicate (N = 80, Poisson regression, residual df = 76, Z stat = −2.68, and P = 0.007). This result is consistent with the finding in the three replication studies that experts could reliably predict which studies would replicate (*11*). This relationship is not solely explained by a priori measures of replicability, such as the statistical power of the original paper. Although replicable publications had higher power (N = 79, Poisson regression, residual df = 75, Z stat = 2.49, and

$P = 0.013$), the relationship between market prices and replicability is still significant, controlling for power, as is the relationship between market prices and citations.

The above results are persistent over time. Yearly citation counts reveal a pronounced gap between papers that replicated and those that did not, as shown in Fig. 3. On average, papers that failed to replicate are cited almost 16 times more per year [random effects Poisson regression, Z stat = −2.84, and $P = 0.005$; column (2) of Table 1]. This difference of 16 citations more per year can be benchmarked against the 5-year impact factor of the journal in which the original studies were published, which measures the citations of papers published in the previous 5 years. In 2016, the 5-year impact factor of *Nature* and *Science* was 44 and 38, respectively, meaning the papers they published in the same time period as the original studies were cited, on average, 38 to 44 times per year. For the two top economics journals considered in (*6*), the impact factor was between 6 and 10, and for the three top psychology journals included in (*5*), it was between 3 and 6. This suggests that the gap in citations is substantial.

The citation gap remains even after the publication of the replication projects. Both results are persistent across several specifications, as we show in the Supplementary Materials using specification curves.

The impact of citations of nonreplicable publications

Understanding the relevance of citations of nonreplicable publications is important. We refer to each citation of the papers that were included in the replication projects as a "citing paper." Do citing papers of nonreplicable and replicable publications in (*5*–*7*) have differential impacts on the field? To measure impact, we consider three metrics: (i) how often citing papers are themselves cited (excluding the replication projects themselves); (ii) whether the citing papers are themselves published in a journal that is included in the Journal Citation Reports (JCR) database, the most comprehensive source of citation data available; and (iii) what is the impact factor of the journals in which citing papers are published. Overall, the data contain 20,252 citing papers.

Figure 4 shows the distribution of citations that citing papers themselves have, for each replication project, separating citing papers that cite a nonreplicable publication from those that cite a replicable one. Papers citing nonreplicable publications are cited 25.6 times, whereas papers citing replicable publications are cited 23.7 times. This difference is not significant (N = 20,252, Poisson regression, residual df = 20,247, Z stat = −0.55, and $P = 0.585$). Detailed regression results are shown in the Supplementary Materials.

The quality of citing papers can also be reflected through journal impact factors. To examine whether the quality of citing papers of nonreplicable and replicable publications differs, we examine whether citing papers of nonreplicable publications are more likely to be published in journals with an impact factor on JCR. Presumably, citing papers of higher quality would be more likely to be published in journals within the JCR database and have a higher average impact factor.

Figure 5 shows that citing papers of replicable publications are more likely to be published in a journal that is in the JCR database. On average, the difference is 6.1 percentage points (N = 20,252, Poisson regression, residual df = 20,247, Z stat = 2.43, and $P = 0.015$). The difference is particularly strong for papers citing papers replicated in the *Nature/Science* and psychology replication projects. However, conditional on being published, citing papers of replicable publications are not published in journals with a higher impact factor (N = 7434, Poisson regression, residual df = 7429, Z stat = −0.36, and $P = 0.722$). Overall, we find a similar impact between the papers

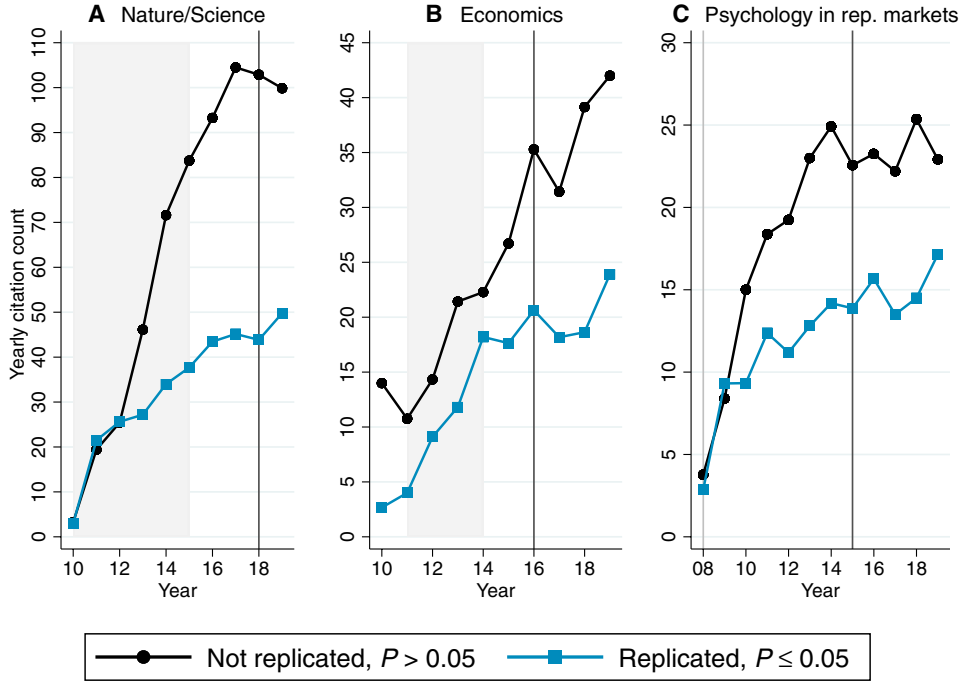

Fig. 3. Yearly citation count by replicability. The average yearly citation count per year for studies that were not replicated (according to *P* value of the replication) in each replication study [(**A**) for *Nature/Science*, (**B**) for Economics, and (**C**) for Psychology papers in replication markets] and for those that were replicated. The light gray area shows the year(s) in which the original studies were published, and the dark line shows the year in which the replication study was published.

Table 1. Yearly citation count and replicability. This table shows the results of a difference-in-differences Poisson regression on citations per year, as downloaded for each replicated paper through Publish or Perish in March of 2020, using a random-effects estimator. The variable Replicated is an indicator variable that takes a value of 1 if the replication study found a significant effect in the same direction as the original study. The variable After publication of replication is an indicator variable that takes a value of 1 for the years after the replication study was published. All regressions include replication project fixed effects. Column (2) also includes the following characteristics of the original study: length of the paper (number of pages), number of authors, share of male authors, whether the experiment was conducted in the United States, whether it was conducted in English, and whether it was conducted online. Standard errors are reported in parentheses. The data source for the citation counts is the software Publish or Perish in March of 2020 (*23*).

	(1)	(2)
	Citations per year	
Replicated ($P < 0.05$)	−15.040***	−15.964***
	(5.341)	(5.631)
After publication of replication	−9.065***	−9.025***
	(1.167)	(1.147)
Replicated × After publication of replication	1.236	1.229
	(0.943)	(0.938)
Years since publication	3.522***	3.506***
	(0.361)	(0.353)
Replication project fixed effects	Yes	Yes
Paper characteristics	No	Yes
Number of papers	80	80
Observations	714	714

***$P < 0.01$.

that cite nonreplicable publications and those that cite replicable publications.

Persistence of citation gap is not explained by negative citations

A driver of citations of nonreplicable publications could be papers written at a later point in time that cite the failed replication. We analyzed how nonreplicable papers are cited after the publication of the replication project. We included the eight *Nature/Science* papers that failed to replicate and examined their citations in 2019 ($N = 798$). We also included the seven economics papers and their citations between 2017 and 2019 ($N = 798$) and 19 out 25 psychology papers (on the basis of a random draw of a subset of the most cited papers) and their citations between 2016 and 2019 ($N = 865$). Of these citing papers, 83% were in English, accessible, and provided a citation in the text to the relevant paper (see details in the Supplementary Materials).

Overall, we find that only 12% of citations after the publication of the replication project acknowledge the replication failure. Within the eight *Nature/Science* papers that failed to replicate, 15% of citations in 2019 acknowledge a replication failure of the original result. Within the economics papers that failed to replicate, 9% of new citations make this acknowledgment (2% in 2017, 14% in 2018, and 9% in 2019). Within the psychology papers that failed to replicate, 12% cite the failure to replicate in the replication project or elsewhere (10% in 2016, 8% in 2017, 21% in 2018, and 7% in 2019).

A possible reason nonreplicable work is cited more is that it is focused on topics about which only a few papers are published. Such citations could nevertheless mention the failure to replicate or the weakness of existing evidence. We do not find evidence supporting this conjecture. Also, note that we cannot test the conjecture that papers are accepted because they are interesting, and hence, our results do not provide evidence on causal relations. We conclude that the gap in citations between replicable and nonreplicable publications is not driven by new papers citing the replication failure.

DISCUSSION

Why are papers that failed to replicate cited more? A possible answer is that the review team may face a trade-off. Although they expect some results to be less robust than others, as shown in the predictions of experts, they are willing to accept this lower expected reliability of the results in some cases. As a result, when the paper is more interesting, the review team may apply lower standards regarding its reproducibility.

A recent book (*18*) suggests that some papers create "hype" using exaggerated and inaccurate claims regarding their findings. According to Richie's argument, the pressure to receive grants and publish favors "…showy and ostentatious findings over workhouse studies that only add small pieces to our knowledge" [(*18*), p. 148]. These studies are also more likely to receive media coverage and become famous. Related to our findings, this exposure may make the papers more likely to be cited. The book also presents evidence that the effect of the hype lingers even after a study is discredited.

Understanding this trade-off is important because it can partially explain the source of the replication crisis in social sciences. It could also help in developing policies to reduce the probability that nonreplicable papers will be accepted for publication. For example, if the results are due to the editor making a trade-off between interesting and reliable results, one way to reduce the occurrence of such incidents is to increase the cost of publishing problematic data, for example, by publishing the name of the editor in the manuscript and going back to them for comment about the editorial process in case the results fail to replicate. A recent example of a move in this direction is the recent retraction of (*19*) from *Psychological Science* (*20*) due to unreliable data. Related to this point, note that when papers are retracted, they are cited significantly less often (*21*).

Another editorial policy that may help in reducing the incentive to publish only remarkable data is registered reports, where the review is conducted before data collection. This process ensures that only the study design (rather than the results being interesting) influences the review process.

We chose to base our analysis on the three replication studies (*5*–*7*) because these projects had objective selection criteria. These objective selection criteria are in contrast to, for example, the Many Labs projects (*8*, *9*), in which researchers selected the papers that were replicated, leading to important selection concerns. The relationship between the P value of the replication and the citation

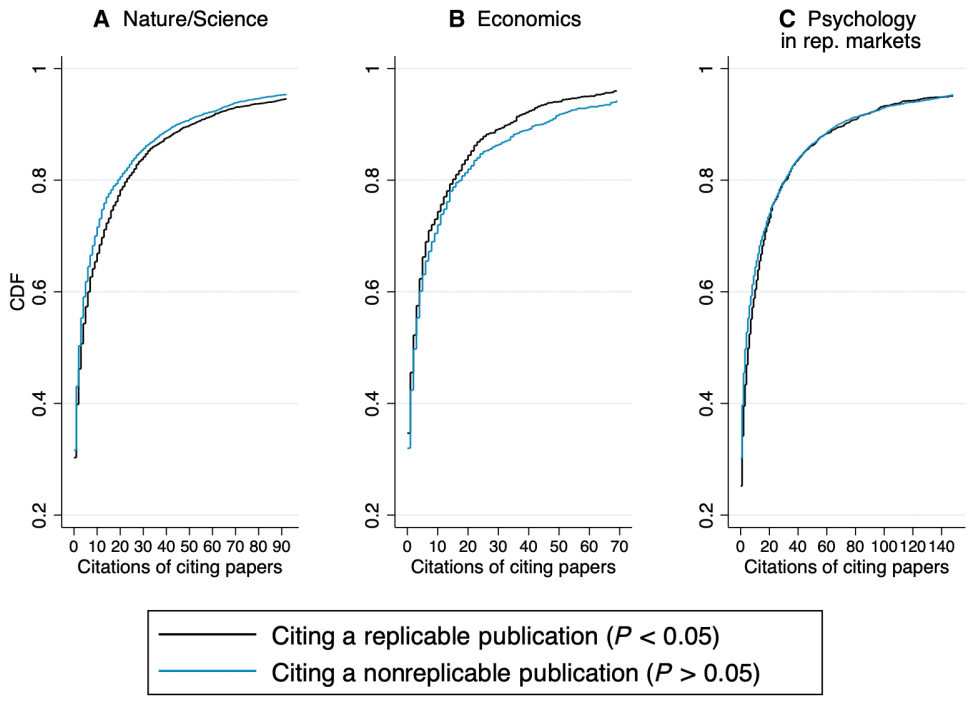

Fig. 4. Distribution of citations that papers citing a replicable versus nonreplicable publication receive. This figure shows the CDF of citations of citing papers, separated by whether the citing paper cited a replicable or a nonreplicable publication for each replication study [(**A**) for *Nature/Science*, (**B**) for Economics, and (**C**) for Psychology papers in replication markets]. The CDF displays 95% of the distribution (to more clearly distinguish the distributions, we exclude the upper 5% of citing papers in terms of citations as their distributions are highly skewed). The replication projects are excluded from the sample of citing papers.

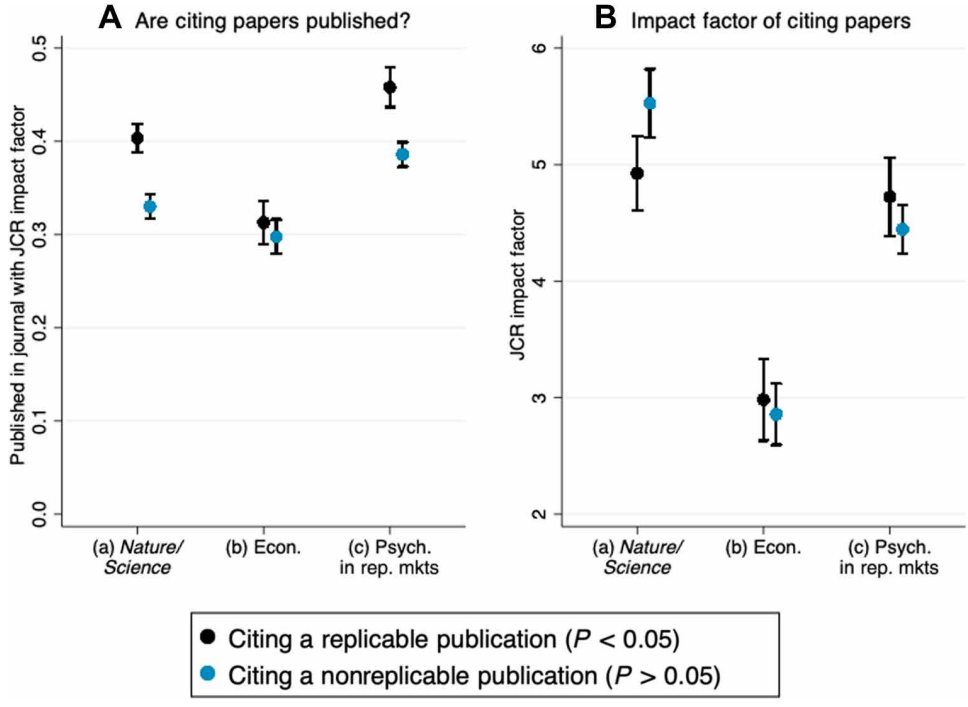

Fig. 5. Likelihood of publication and impact factor of papers citing a replicable versus nonreplicable publication. This figure shows the fraction of citing papers that are published in a journal that is listed in the JCR database (**A**) and the average JCR impact factor (**B**), separated by whether the citing paper cited a replicable or a nonreplicable publication. Bars show 95% confidence intervals. The replication projects are excluded from the sample of citing papers.

count of the original study, combining all replication projects without covariates, is slightly negative (*15*).

As always, one should be cautious when generalizing from data. We cannot tell, for example, whether our findings will be true for comparable papers in other fields or other journals. For example, only focusing on the replication studies in psychology (*5*) and including all replications regardless of whether replication markets were run reveals that there may be no difference in citations and citations of citing papers (*22*). Still, together, these findings reveal that the failure to replicate does not lead to fewer citations, as one may expect.

Replication projects also typically replicated only one study out of several results presented in the original paper. For instance, in the case of the economics replications, only the first result of each study was replicated, even if this was not the main point of the paper. The replication may also fail to reflect exactly how the original study was conducted. Thus, the results of the replication are a noisy measure of the replicability of the entire paper. In addition, we do not observe studies that journals rejected. Hence, our findings are based on the set of selected papers that ended up being published in the respective prestigious journals.

Another interesting question that should be investigated in future research is how long the effect we found will hold. For example, will the field eventually internalize a paper's failure to replicate and will its impact be reduced? We can reasonably assume that the attention in the literature to the replication crisis will improve practices and result in reduced failed replications. We hope that our findings will contribute to this change.

MATERIALS AND METHODS

This paper focuses on the replicability of studies included in three large-scale replication projects (*5*–*7*). The inclusion criteria for studies in these projects were all experimental papers published in a set of leading journals within a given period, eliminating selection concerns and making the replicated studies as comparable as possible. The analysis focuses on papers for which a prediction market was run, including all replications in Camerer *et al.* (*6*, *7*), and 41 of the 100 replications in (*5*). Results are qualitatively similar if the dataset is extended to include all 96 replications in (*5*) of papers that had a significant effect originally, as shown in the Supplementary Materials. Focusing only on the psychology replication project, the gap in citations of replicable versus nonreplicable papers is small and not significant, as shown in (*22*). We obtained yearly citation counts in March of 2020, using the software Publish or Perish (*23*).

Following the literature (*5*–*7*), the main measure of replicability that we use is whether the *P* value of the replication study is below 0.05 in a two-sided test and with an effect in the same direction as the original one. To remain consistent with the 0.05 cutoff, we also refer to our results as significant if their *P* value is below 0.05. The relationship between replicability and citation also remains significant considering a 0.005 cutoff, as suggested in (*24*). We also consider the relative effect size of the replication, defined as the ratio of the effect size in the replication over the effect size in the original study. The relative effect size ranges between −0.90 and 2.38 for 79 of the 80 studies. We find one outlier, with an effect size of 22.82, which we exclude from the analyses. As a measure of experts' beliefs, we use prices in the prediction markets. These prices reflect the belief that the replication study will replicate the original paper (*11*).

As shown in the Supplementary Materials, the distribution of citation counts is highly right-skewed. We hence use Poisson regression models for the main specification in the paper. Poisson regression models are particularly well suited to model count data, as is the case with citation counts (*25*). These models are estimated via maximum likelihood. We consider up to 54 different specifications of the model and show the results using specification curves in the Supplementary Materials. The conclusions remain qualitatively similar when we include characteristics of the paper, such as the number of authors and the field (social/cognitive psychology and economics) to which the paper contributes (*15*, *16*). Related work leveraged machine-learning methods to predict the replicability of scientific papers, using the 96 papers included in the psychology replication project to train their model (*15*, *21*, *26*). In addition to study characteristics, they used text analysis and do not find that words such as "remarkable" or "unexpected" predicted the replicability out of sample. A possible reason for this is that the authors of the papers may not specifically say their results are remarkable or unexpected, even if they are in the eyes of others in the profession.

To document the impact of citations, which we refer to as citing papers, we use the information provided by Publish or Perish. The software tracks the number of times that citing papers have been cited themselves. It also documents which journals the papers were published in, if at all. After careful data cleaning, as explained in the Supplementary Materials, we created a dataset with the journal names in which citing papers were published and matched them with a database of JCR impact factors from the Web of Science group (Clarivate Analytics). The JCR database is the most comprehensive source of citation data available. It generates impact factors for more than 12,000 journals and conference proceedings in more than 80 countries (*27*). In the analyses, we focus on the impact factor of the journal in 2019, though results remain qualitatively similar if we use the 5-year impact factor of the journal.

We then carefully study the type of citations that arise after the publication of the replication projects. We downloaded available citing papers of failed replications that were published after the publication of the replication project (in 2016 or later for the psychology replication project, in 2017 or later for the economics replication project, and in 2019 for the *Nature/Science* replication project). As documented in detail in the Supplementary Materials, we then skimmed each citing paper and checked the places in each citing paper in which the failed replication was cited. If the failure to replicate or "mixed" evidence regarding the findings of the failed replication were cited, we considered the paper as a negative citation, in which the failure to replicate (or at least the existence of opposing results) was acknowledged.

For the economics and *Nature/Science* replication projects, we considered all failed replications. For the psychology replication project, we classified the citing papers of 19 of 25 failed replications. Because of the number of citing papers, we did not do this exercise for all failed replications: We included all citing papers of those failed replications that were cited less than 100 times since 2015 (17 papers) and randomly drew two citing papers from those that were cited more than 100 times (8 papers). Further details of the procedures and the number of citing papers we were able to classify for each failed replication are provided in the Supplementary Materials.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/21/eabd1705/DC1

REFERENCES AND NOTES

1. B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. L. Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. Vanden Bos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni, Promoting an open research culture. *Science* **348**, 1422–1425 (2015).
2. J. Simmons, L. Nelson, U. Simonsohn, False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
3. E. Vivalt, Specification searching and significance inflaction across time, methods and disciplines. *Oxford B. Econ. Stat.* **81**, 797–816 (2019).
4. A. Brodeur, N. Cook, A. Heyes, Methods matter: P-hacking and publication bias in causal analysis in economics. *Am. Econ. Rev.* **110**, 3634–3660 (2020).
5. Open Science Collaboration, Estimating the reproducibility of psychological science. *Science* **349**, aa4716 (2015).
6. C. F. Camerer, A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, H. Wu, Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).
7. C. F. Camerer, A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, A. Altmejd, N. Buttrick, T. Chan, Y. Chen, E. Forsell, A. Gampa, E. Heikensten, L. Hummer, T. Imai, S. Isaksson, D. Manfredi, J. Rose, E.-J. Wagenmakers, H. Wu, Evaluating replicability of social science experiments published in *Nature* and *Science* between 2010 and 2015. *Nat. Human Behav.* **2**, 637–644 (2018).
8. R. A. Klein, K. A. Ratliff, M. Vianello, R. B. Adams Jr., Š. Bahník, M. J. Bernstein, K. Bocian, M. J. Brandt, B. Brooks, C. C. Brumbaugh, Z. Cemalcilar, J. Chandler, W. Cheong, W. E. Davis, T. Devos, M. Eisner, N. Frankowska, D. Furrow, E. M. Galliani, F. Hasselman, J. A. Hicks, J. F. Hovermale, S. J. Hunt, J. R. Huntsinger, H. IJzerman, M.-S. John, J. A. Joy-Gaba, H. B. Kappes, L. E. Krueger, J. Kurtz, C. A. Levitan, R. K. Mallett, W. L. Morris, A. J. Nelson, J. A. Nier, G. Packard, R. Pilati, A. M. Rutchick, K. Schmidt, J. L. Skorinko, R. Smith, T. G. Steiner, J. Storbeck, L. M. Van Swol, D. Thompson, A. E. van 't Veer, L. A. Vaughn, M. Vranka, A. L. Wichman, J. A. Woodzicka, B. A. Nosek, Investigating variation in replicability: A "Many Labs" replication project. *Soc. Psychol.* **45**, 142–152 (2014).
9. R. A. Klein, M. Vianello, F. Hasselman, B. G. Adams, R. B. Adams Jr., S. Alper, M. Aveyard, J. R. Axt, M. T. Babalola, Š. Bahník, R. Batra, M. Berkics, M. J. Bernstein, D. R. Berry, O. Bialobrzeska, E. D. Binan, K. Bocian, M. J. Brandt, R. Busching, A. C. Rédei, H. Cai, F. Cambier, K. Cantarero, C. L. Carmichael, F. Ceric, J. Chandler, J.-H. Chang, A. Chatard, E. E. Chen, W. Cheong, D. C. Cicero, S. Coen, J. A. Coleman, B. Collisson, M. A. Conway, K. S. Corker, P. G. Curran, F. Cushman, Z. K. Dagona, I. Dalgar, A. D. Rosa, W. E. Davis, M. de Bruijn, L. De Schutter, T. Devos, M. de Vries, C. Doğulu, N. Dozo, K. N. Dukes, Y. Dunham, K. Durrheim, C. R. Ebersole, J. E. Edlund, A. Eller, A. S. English, C. Finck, N. Frankowska, M.-Á. Freyre, M. Friedman, E. M. Galliani, J. C. Gandi, T. Ghoshal, S. R. Giessner, T. Gill, T. Gnambs, Á. Gómez, R. González, J. Graham, J. E. Grahe, I. Grahek, E. G. T. Green, K. Hai, M. Haigh, E. L. Haines, M. P. Hall, M. E. Heffernan, J. A. Hicks, P. Houdek, J. R. Huntsinger, H. P. Huynh, H. IJzerman, Y. Inbar, Å. H. Innes-Ker, W. Jiménez-Leal, M.-S. John, J. A. Joy-Gaba, R. G. Kamiloğlu, H. B. Kappes, S. Karabati, H. Karick, V. N. Keller, A. Kende, N. Kervyn, G. Knežević, C. Kovacs, L. E. Krueger, G. Kurapov, J. Kurtz, D. Lakens, L. B. Lazarević, C. A. Levitan, N. A. Lewis Jr., S. Lins, N. P. Lipsey, J. E. Losee, E. Maassen, A. T. Maitner, W. Malingumu, R. K. Mallett, S. A. Marotta, J. Međedović, F. Mena-Pacheco, T. L. Milfont, W. L. Morris, S. C. Murphy, A. Myachykov, N. Neave, K. Neijenhuijs, A. J. Nelson, F. Neto, A. L. Nichols, A. Ocampo, S. L. O'Donnell, H. Oikawa, M. Oikawa, E. Ong, G. Orosz, M. Osowiecka, G. Packard, R. Pérez-Sánchez, B. Petrović, R. Pilati, B. Pinter, L. Podesta, G. Pogge, M. M. H. Pollmann, A. M. Rutchick, P. Saavedra, A. K. Saeri, E. Salomon, K. Schmidt, F. D. Schönbrodt, M. B. Sekerdej, D. Sirlopú, J. L. M. Skorinko, M. A. Smith, V. Smith-Castro, K. C. H. J. Smolders, A. Sobkow, W. Sowden, P. Spachtholz, M. Srivastava, T. G. Steiner, J. Stouten, C. N. H. Street, O. K. Sundfelt, S. Szeto, E. Szumowska, A. C. W. Tang, N. Tanzer, M. J. Tear, J. Theriault, M. Thomae, D. Torres, J. Traczyk, J. M. Tybur, A. Ujhelyi, R. C. M. van Aert, M. A. L. M. van Assen, M. van der Hulst, P. A. M. van Lange, A. E. van 't Veer, A. Vásquez-Echeverría, L. A. Vaughn, A. Vázquez, L. D. Vega, C. Verniers, M. Verschoor, I. P. J. Voermans, M. A. Vranka, C. Welch, A. L. Wichman, L. A. Williams, M. Wood, J. A. Woodzicka, M. K. Wronska, L. Young, J. M. Zelenski, Z. Zhijia, B. A. Nosek, Many Labs 2: Investigating variation in replicability across sample and setting. *Adv. Methods Pract. Psychol. Sci.* **1**, 443–490 (2018).
10. C. R. Ebersole, O. E. Atherton, A. L. Belanger, H. M. Skulborstad, J. M. Allen, J. B. Banks, E. Baranski, M. J. Bernstein, D. B. V. Bonfiglio, L. Boucher, E. R. Brown, N. I. Budiman, A. H. Cairo, C. A. Capaldi, C. R. Chartier, J. M. Chung, D. C. Cicero, J. A. Coleman, J. G. Conway, W. E. Davis, T. Devos, M. M. Fletcher, K. German, J. E. Grahe, A. D. Hermann, J. A. Hicks, N. Honeycutt, B. Humphrey, M. Janus, D. J. Johnson, J. A. Joy-Gaba, H. Juzeler, A. Keres, D. Kinney, J. Kirshenbaum, R. A. Klein, R. E. Lucas, C. J. N. Lustgraaf, D. Martin, M. Menon, M. Metzger, J. M. Moloney, P. J. Morse, R. Prislin, T. Razza, D. E. Re, N. O. Rule, D. F. Sacco, K. Sauerberger, E. Shrider, M. Shultz, C. Siemsen, K. Sobocko, R. W. Sternglanz, A. Summerville, K. O. Tskhay, Z. Allen, L. A. Vaughn, R. J. Walker, A. Weinberg, J. P. Wilson, J. H. Wirth, J. Wortman, B. A. Nosek, Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *J. Exp. Soc. Psychol.* **67**, 68–82 (2016).
11. A. Dreber, T. Pfeiffer, J. Almenberg, S. Isaksson, B. Wilson, Y. Chen, B. Nosek, M. Johannesson, Using prediction markets to estimate the reproducibility of scientific research. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15343–15347 (2015).
12. K. Siler, K. Lee, L. Bero, Measuring the effectiveness of scientific gatekeeping. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 360–365 (2014).
13. D. Card, S. DellaVigna, What do editors maximize? Evidence from four economics journals. *Rev. Econ. Stat.* **102**, 195–217 (2020).
14. E. Forsell, D. Viganola, T. Pfeiffer, J. Almenberg, B. Wilson, Y. Chen, B. Nosek, M. Johannesson, A. Dreber, Predicting replication outcomes in the Many Labs 2 study. *J. Econ. Psychol.* **75**, 102117 (2019).
15. A. Altmejd, A. Dreber, E. Forsell, J. Huber, T. Imai, M. Johannesson, M. Kirchler, G. Nave, C. Camerer, Predicting the replicability of social science lab experiments. *PLOS ONE* **14**, e0225826 (2019).
16. Y. Inbar, Association between contextual dependence and replicability in psychology may be spurious. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4933–E4934 (2016).
17. U. Simmonsohn, J. Simmons, L. D. Nelson, Specification curve analysis. *Nat. Human Behav.* **4**, 1208–1214 (2020).
18. S. Richie, *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth* (Metropolitan Books, 2020).
19. C. J. Clark, B. M. Winegard, J. Beardslee, R. F. Baumeister, A. F. Shariff, RETRACTED: Declines in religiosity predict increases in violent crime—but not among countries with relatively high average IQ. *Psychol. Sci.* **31**, 170–183 (2020).
20. Retraction Watch (June 27, 2020), Editors in chief past and present apologize for publishing article that feed[s] into racist narratives retrieved from https://retractionwatch.com/2020/06/27/editors-in-chief-past-and-present-apologize-for-publishing-article-that-feeds-into-racist-narratives/.
21. S. F. Lu, G. Z. Jin, B. Uzzi, B. Jones, The Retraction Penalty: Evidence from the Web of Science. *Sci. Rep.* **3**, 3146 (2013).
22. Y. Yang, W. Youyou, B. Uzzi, Estimating the deep replicability of scientific findings using human and artificial intelligence. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10762–10768 (2020).
23. A. W. Harzing (2007), Publish or Perish, available from https://harzing.com/resources/publish-or-perish.
24. D. Benjamin, J. O. Berger, M. Johannesson, B. A. Nosek, E.-J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer, D. Cesarini, C. D. Chambers, M. Clyde, T. D. Cook, P. De Boeck, Z. Dienes, A. Dreber, K. Easwaran, C. Efferson, E. Fehr, F. Fidler, A. P. Field, M. Forster, E. I. George, R. Gonzalez, S. Goodman, E. Green, D. P. Green, A. G. Greenwald, J. D. Hadfield, L. V. Hedges, L. Held, T. H. Ho, H. Hoijtink, D. J. Hruschka, K. Imai, G. Imbens, J. P. A. Ioannidis, M. Jeon, J. H. Jones, M. Kirchler, D. Laibson, J. List, R. Little, A. Lupia, E. Machery, S. E. Maxwell, M. M. Carthy, D. A. Moore, S. L. Morgan, M. Munafó, S. Nakagawa, B. Nyhan, T. H. Parker, L. Pericchi, M. Perugini, J. Rouder, J. Rousseau, V. Savalei, F. D. Schönbrodt, T. Sellke, B. Sinclair, D. Tingley, T. Van Zandt, S. Vazire, D. J. Watts, C. Winship, R. L. Wolpert, Y. Xie, C. Young, J. Zinman, V. E. Johnson, Redefine statistical significance. *Nat. Human Behav.* **2**, 6–10 (2018).
25. A. C. Cameron, P. K. Trivedi, *Regression Analysis of Count Data* (Cambridge Univ. Press, ed. 2, 1998).
26. S. Pawel, L. Held, Probabilistic forecasting of replication studies. *PLOS ONE* **15**, e0231416 (2020).
27. Measuring your Research Impact: Journal Citation Reports (April 9, 2020); https://ucsd.libguides.com/ResearchImpact/JCR.
28. J. A. Nelder, R. W. M. Wedderburn, Generalized linear models. *J. R. Stat. Soc. A* **135**, 370–384 (1972).

Acknowledgments: We thank C. Hastings, E. Amozegar, and W. W.-Y. Wang for excellent research assistance. **Funding:** The authors have no funding to declare. **Author contributions:** Both authors contributed equally to this work. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The raw data and STATA codes used for analyses are posted on OSF (https://osf.io/dx3pk/?view_only=2f5b2c56bb6f4f609f1426848cec78b2) as supplementary information of the published article.

Submitted 4 June 2020
Accepted 1 April 2021
Published 21 May 2021
10.1126/sciadv.abd1705

Citation: M. Serra-Garcia, U. Gneezy, Nonreplicable publications are cited more than replicable ones. *Sci. Adv.* **7**, eabd1705 (2021).

Nonreplicable publications are cited more than replicable ones

Marta Serra-GarciaUri Gneezy

Sci. Adv., 7 (21), eabd1705. • DOI: 10.1126/sciadv.abd1705