nonradiatively [calculation in the supplementary materials (21)] in a perovskite film in air on glass. This value relates to a photon recycling–assisted average excitation travel distance of 20 µm (fig. S18). The average travel distance could be enhanced at larger charge densities (for example, under high fluences) and can reach values beyond 50 µm.

In terms of e^--h^+ transport, our results suggest that the average distance a charge carrier can travel in a perovskite is not limited by the charge-carrier diffusion length, for as long as recombination is radiative and the photon stays in the film, the e^--h^+ pair can be regenerated and can propagate over large distances. This process creates a distinction between extraction and charge diffusion lengths and allows us to solve the existing contradiction of reported high recombination rates and long diffusion lengths.

What are the implications of the observations presented here for standard thin-film perovskite solar cells (3, 6)? The thin-film samples from our work provide valuable model systems for these structures. Using the model and parameters developed above, we estimate that, under open-circuit conditions, in a device with a thickness of 350 nm and nonquenching electrodes, recycling produces a doubling of the internal photon density under 1-sun illumination. These effects can be enhanced further by minimizing nonradiative decay channels and being subjected to higher fluences, such as in solar concentrators, where high bimolecular recombination rates dominate. In the ideal case of unity PLQE and a perfect back mirror, photon recycling can produce internal photon densities up to 25 suns ($4n^2$ with $n = 2.5$) (31) in perovskite solar cells under open-circuit conditions. Photon management, such as the use of highly reflective back mirrors to minimize photonic losses and texturing of the top surface, offers promising approaches for using photon recycling to improve photoconversion efficiencies of perovskite solar cells toward the Shockley-Queisser limit. Higher photon densities lead to higher internal luminescence and a build-up of excited charges, which increase the split of quasi-Fermi levels and enhance the achievable open-circuit voltage in a solar cell.

REFERENCES AND NOTES

1. A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka, *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).
2. J.-H. Im, C.-R. Lee, J.-W. Lee, S.-W. Park, N.-G. Park, *Nanoscale* **3**, 4088–4093 (2011).
3. M. M. Lee, J. Teuscher, T. Miyasaka, T. N. Murakami, H. J. Snaith, *Science* **338**, 643–647 (2012).
4. M. Liu, M. B. Johnston, H. J. Snaith, *Nature* **501**, 395–398 (2013).
5. H.-S. Kim et al., *Sci. Rep.* **2**, 591 (2012).
6. N. J. Jeon et al., *Nature* **517**, 476–480 (2015).
7. J. Burschka et al., *Nature* **499**, 316–319 (2013).
8. S. De Wolf et al., *J. Phys. Chem. Lett.* **5**, 1035–1039 (2014).
9. C. Wehrenfennig, G. E. Eperon, M. B. Johnston, H. J. Snaith, L. M. Herz, *Adv. Mater.* **26**, 1584–1589 (2014).
10. F. Deschler et al., *J. Phys. Chem. Lett.* **5**, 1421–1426 (2014).
11. Q. Dong et al., *Science* **347**, 967–970 (2015).
12. D. Shi et al., *Science* **347**, 519–522 (2015).
13. G. Xing et al., *Science* **342**, 344–347 (2013).
14. S. D. Stranks et al., *Science* **342**, 341–344 (2013).
15. E. Yablonovitch, O. D. Miller, S. R. Kurtz, in *Conference Record of the 38th IEEE Photovoltaic Specialists Conference*, Austin, TX, 3 to 8 June 2012 (IEEE, 2012), pp. 1556–1559.
16. L. S. Mattos et al., in *Conference Record of the 38th IEEE Photovoltaic Specialists Conference*, Austin, TX, 3 to 8 June 2012 (IEEE, 2012), pp. 3187–3190.
17. O. D. Miller, E. Yablonovitch, S. R. Kurtz, *IEEE J. Photovoltaics* **2**, 303–311 (2012).
18. E. Dupont, H. C. Liu, M. Buchanan, S. Chiu, M. Gao, *Appl. Phys. Lett.* **76**, 4 (2000).
19. A. Sadhanala et al., *J. Phys. Chem. Lett.* **5**, 2501–2505 (2014).
20. C. Wehrenfennig, M. Liu, H. J. Snaith, M. B. Johnston, L. M. Herz, *J. Phys. Chem. Lett.* **5**, 1300–1306 (2014).
21. Full details can be found in the supplementary materials on *Science* Online.
22. I. Suárez, E. J. Juárez-Pérez, J. Bisquert, I. Mora-Seró, J. P. Martínez-Pastor, *Adv. Mater.* **27**, 6157–6162 (2015).
23. W. Zhang et al., *Nat. Commun.* **6**, 6142 (2015).
24. Y. Yamada, T. Nakamura, M. Endo, A. Wakamiya, Y. Kanemitsu, *J. Am. Chem. Soc.* **136**, 11610–11613 (2014).
25. J. L. Balenzategui, A. Martí, *Sol. Energy Mater. Sol. Cells* **90**, 1068–1088 (2006).
26. S. M. Durbin, J. L. Gray, *IEEE Trans. Electron. Dev.* **41**, 239–245 (1994).
27. R. Graaff, J. J. Ten Bosch, *Opt. Lett.* **25**, 43–45 (2000).
28. M. B. Johnston, L. M. Herz, *Acc. Chem. Res.* **49**, 146–154 (2016).
29. R. K. Ahrenkiel et al., *Appl. Phys. Lett.* **55**, 1088–1090 (1989).
30. P. Renaud, F. Raymond, B. Bensaïd, C. Vèrié, *J. Appl. Phys.* **71**, 1907–1913 (1992).
31. Q. Lin, A. Armin, R. C. R. Nagiri, P. L. Burn, P. Meredith, *Nat. Photonics* **9**, 106–112 (2014).

ECONOMICS

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer,[1]*† Anna Dreber,[2]† Eskil Forsell,[2]† Teck-Hua Ho,[3,4]† Jürgen Huber,[5]† Magnus Johannesson,[2]† Michael Kirchler,[5,6]† Johan Almenberg,[7] Adam Altmejd,[2] Taizan Chan,[8] Emma Heikensten,[2] Felix Holzmeister,[5] Taisuke Imai,[1] Siri Isaksson,[2] Gideon Nave,[1] Thomas Pfeiffer,[9,10] Michael Razen,[5] Hang Wu[4]

The replicability of some scientific findings has recently been called into question. To contribute data about replicability in economics, we replicated 18 studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. All of these replications followed predefined analysis plans that were made publicly available beforehand, and they all have a statistical power of at least 90% to detect the original effect size at the 5% significance level. We found a significant effect in the same direction as in the original study for 11 replications (61%); on average, the replicated effect size is 66% of the original. The replicability rate varies between 67% and 78% for four additional replicability indicators, including a prediction market measure of peer beliefs.

The deepest trust in scientific knowledge comes from the ability to replicate empirical findings directly and independently. Although direct replication is widely applauded (1), it is rarely carried out in empirical social science. Replication is now more important than ever, because the quality of results has been questioned in many fields, such as medicine (2–5), neuroscience (6), and genetics (7, 8). In economics, concerns about inflated findings in empirical (9) and experimental analyses (10, 11) have also been raised. In the social sciences, psychology has been the most active in both self-diagnosing the forces that create "false positives" and conducting direct replications (12–15). Several high-profile replication failures (16, 17) quickly led to changes in journal publication practices (18). The recent Reproducibility Project: Psychology (RPP) replicated 100 original studies published in three top journals in psychology. The vast majority (97) of the original studies reported "positive findings," but in the replications, the RPP only found a significant effect in the same direction for 36% of these studies (19).

In this report, we provide insights into the replicability of laboratory experiments in economics. Our sample consists of all 18 between-subject laboratory experimental papers published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. The most important statistically significant finding,

as emphasized by the authors of each paper, was chosen for replication (see section 1 of the supplementary materials and tables S1 and S2). We used replication sample sizes with at least 90% power (mean = 92%; median = 91%) to detect the original effect size at the 5% significance level. All of the replication and analysis plans were made public on the project website (supplementary materials, section 1) and were also sent to the original authors for verification.

There are different ways of assessing replication, with no universally agreed-upon standard of excellence (19–23). We present results for the same replication indicators that were used in the RPP (19). As our first indicator of replication, we used a "significant effect in the same direction as in the original study" [Gelman and Stern (20) discuss the challenges of comparing significance levels across experiments].

The results of the replications are shown in Fig. 1A and table S1. We found a significant effect

[1]Division of Humanities and Social Sciences, California Institute of Technology, 1200 East California Boulevard, MC 228-77, Pasadena, CA 91125, USA. [2]Department of Economics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden. [3]Haas School of Business, University of California–Berkeley, Berkeley, CA 94720-1900, USA. [4]NUS Business School, National University of Singapore 119245. [5]Department of Banking and Finance, University of Innsbruck, Universitätsstrasse 15, 6020 Innsbruck, Austria. [6]Centre for Finance, Department of Economics, University of Göteborg, SE-40530 Göteborg, Sweden. [7]Sveriges Riksbank, SE-103 37 Stockholm, Sweden. [8]Office of the Deputy President (Research and Technology), National University of Singapore, Singapore 119077. [9]New Zealand Institute for Advanced Study, Private Bag 102904, North Shore Mail Centre, Auckland 0745, New Zealand. [10]Wissenschaftskolleg zu Berlin, Institute for Advanced Study, D-14193 Berlin, Germany.
*Corresponding author. E-mail: camerer@hss.caltech.edu
†These authors contributed equally to this work.

in the same direction as in the original study for 11 replications (61.1%). This is considerably lower than the replication rate of 92% (mean power) that would be expected if all original effects were true and accurately estimated (one-sample binomial test, $P < 0.001$).

A complementary method for assessing replicability is to test whether the 95% confidence interval (CI) of the replication effect size includes the original effect size (19) [Cumming (21) discusses the interpretation of CIs for replications]. This is the case in 12 of our replications (66.7%). If we also include the study in which the entire 95% CI exceeds the original effect size, the number of replicable studies increases to 13 (72.2%). An alternative measure, which acknowledges sampling error in both the original study and the replications, is to count how many replicated effects lie in a 95% "prediction interval" (24). This count is higher (83.3%) and increases to 88.9% if we also include the replication whose effect size exceeds the upper bound of the prediction interval (fig. S2 and supplementary materials, section 2).

The mean standardized effect size (correlation coefficient, r) of the replications is 0.279, compared with 0.474 in the original studies (fig. S3). This difference is significant [Wilcoxon signed-rank test; $z = -2.98$, $P = 0.003$, $n = 18$]. The replicated effect sizes tend to be of the same sign as the original ones but not as large. The mean relative effect size of the replications is 65.9%.

The original and replication studies can also be combined in a meta-analytic estimate of the effect size (19). As shown in Fig. 1B, in the meta-analysis, 14 studies (77.8%) have a significant effect in the same direction as in the original study. These results should be interpreted cau-

tiously, because the estimates assume that the results of the original studies do not have publication or reporting biases.

To measure peer beliefs about the replicability of original results, we set up prediction markets before the 18 replications were performed (25). Dreber et al. (26), in a recent study that presented evidence for a subset of the replications in the RPP, proposed the use of prediction markets as an additional replicability indicator. In the prediction market for a particular target study, peers who were likely to be familiar with experimental methods in economics could buy or sell shares whose monetary value depended on whether the target study was replicated (fig. S4 and tables S1 and S2). The prediction markets produce a collective market probability of replication (27) that can be interpreted as a replicability indicator (26). The traders' ($n = 97$) survey beliefs about replicability were also collected before market trading as an additional measure of peer beliefs.

The average prediction market belief is a replication rate of 75.2%, and the average survey belief is 71.1% (Fig. 2, fig. S5, and tables S3 and S4). Both are higher than the observed replication rate of 61.1%, but neither difference is significant (supplementary Materials, section 5). The prediction market beliefs and the survey beliefs are highly correlated, and both are positively correlated with the ranked degree of replication success, although the correlation does not reach significance for the prediction market beliefs (Fig. 2 and fig. S6). Contrary to Dreber et al. (26), prediction market beliefs are not a more accurate indicator of replicability than survey beliefs.

We also tested whether replicability is correlated with two observable characteristics of

Fig. 1. Replication results.
(**A**) Plotted are 95% CIs of replication effect sizes (standardized to correlation coefficients). The standardized effect sizes are normalized so that 1 equals the original effect size (fig. S1 shows a nonnormalized version). Eleven replications have a significant effect in the same direction as in the original study [61.1%; 95% CI = (36.2%, 86.1%)]. The 95% CI of the replication effect size includes the original effect size for 12 replications [66.7%; 95% CI = (42.5%, 90.8%)]; if we also include the study in which the entire 95% CI exceeds the original effect size, this increases to 13 replications [72.2%; 95% CI = (49.3%, 95.1%)]. AER denotes the *American Economic Review* and QJE denotes the *Quarterly Journal of Economics*.
(**B**) Meta-analytic estimates of

effect sizes, combining the original and replication studies. Plotted are 95% CIs of combined effect sizes (standardized to correlation coefficients). The standardized effect sizes are normalized as in (A) (fig. S1 shows a nonnormalized version). Fourteen studies have a significant effect in the same direction as the original study in the meta-analysis [77.8%; 95% CI = (56.5%, 99.1%)].
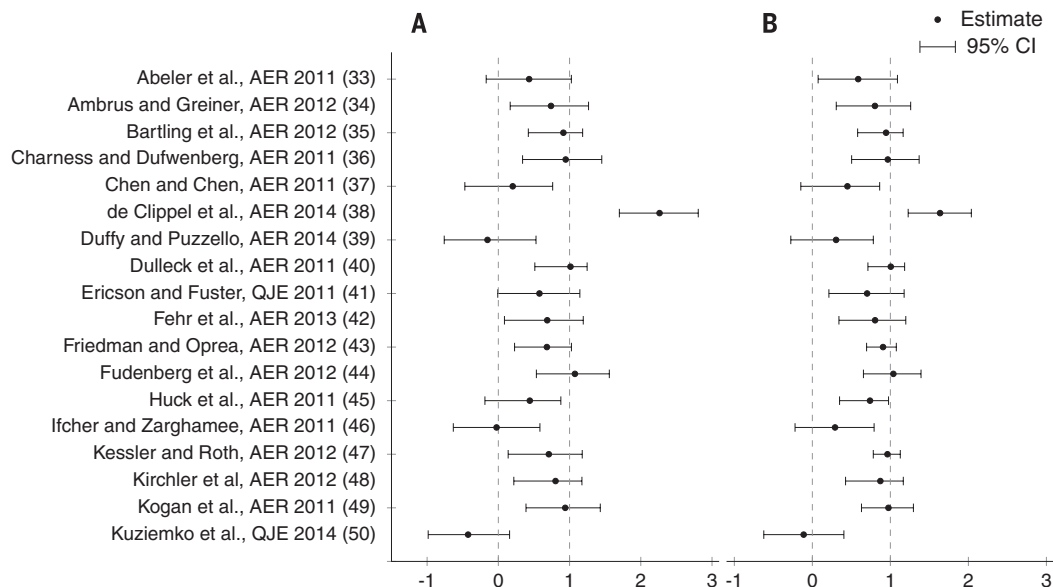
published studies: the *P* value and the sample size (number of participants) of the original study. These two characteristics are likely to be correlated with each other, which is the case for our 18 studies (Spearman correlation coefficient = −0.61, *P* = 0.007, *n* = 18). We expected the replicability to be negatively correlated with the original *P* value and positively correlated with the sample size, because the risk of false positives increases with the original *P* value and decreases with the original sample size (statistical power) (*6*, *11*). The correlations are presented in Fig. 3 and table S5, and the results are in line with our expectations. The correlations are typically around 0.5 in the expected direction and significant. Only one study out of eight with a *P* value <0.01 in the original study was not replicable at the 5% level in the original direction.

We report the first systematic replications of laboratory experiments in economics, with the aim of contributing much-needed data to the larger question of the replicability of empirical findings in all areas of science. The results provide provisional answers to two questions: (i) Are laboratory experiments in economics generally replicable, and (ii) do statistical measures of research quality, including peer beliefs about replicability, help predict which studies will be replicable?

The provisional answer to the first question is that, based on this sample of experiments, replication is generally possible, although there is room for improvement. Eleven out of 18 (61.1%) studies were replicable with *P* < 0.05 in the original direction, and three more studies were relatively close to being replicated (all have sig-

nificant effects in the meta-analysis). Four replications (22.2%) had effect sizes close to zero, somewhat more than the 1.4 replication failures expected by pure chance (given the mean power of 92%). Moreover, the original effect sizes in the studies that we replicated could have been inflated, a phenomenon that could stem from publication bias (*28*). If there is publication bias, our prospective power analyses will have overestimated the replication power.

The answer to the second question is that peer surveys and market beliefs did contain some information about which experiments were more likely to replicate, but sample sizes and *P* values in the original studies were even more strongly correlated with replicability (Fig. 3).

To learn from successes and failures in different scientific fields, it is useful to compare our results with recent results from studies of robustness in experimental psychology and empirical economics. Our results can be compared with the recent RPP project in the psychological sciences (*19*), which was also accompanied by prediction market beliefs and survey beliefs (*26*). All measures of replication success are somewhat higher for the economics experiments than for the sampled psychology experiments (Fig. 4). Peer beliefs in our study are also significantly higher than in the RPP study (Fig. 4). Acknowledging the limits of this two-study comparison, and particularly our small sample of 18 replications, there appears to be some difference in replication success between these fields. However, it is premature to draw strong conclusions about disciplinary differences; other methodological factors potentially could explain why the replication rates differed. For example, in the RPP replications, interaction effects were less likely to be replicable than main or simple effects (*19*).

In economics, several studies have shown that statistical findings from nonexperimental data are not always easy to replicate (*29*). Two studies of macroeconomic findings, reported in the *Journal of Money, Credit and Banking* in 1986 and 2006, respectively found that only 13% and 23% of original results were replicable, even when the data and code were easily accessible (*30*, *31*). An analysis of 50,000 *P* values reported between 2005 and 2011 in three widely cited general economics journals found that *P* values between 0.10 and 0.25 were less common than might be expected. (*32*). However, the frequency of these "missing" *P* values is smaller in laboratory and field experiments. Taken together, these analyses and our replication sample suggest that laboratory experiments are at least as robust, and perhaps more robust, than other kinds of empirical economics.

Two methodological research practices in laboratory experimental economics may contribute to relatively high replication success. First, experimental economists have strong norms about motivating subjects with substantial financial incentives and avoiding the use of deception. These norms make subjects more responsive and may reduce variability in how experiments are performed across different research teams,

Fig. 2. Prediction market and survey beliefs. A plot of prediction market beliefs and survey beliefs, in relation to whether the original result was replicated with *P* < 0.05 in the original direction. The mean prediction market belief in a successful replication is 75.2% [range, 59% to 94%; 95% CI = (69.7%, 80.6%)], and the mean survey belief is 71.1% [range, 54% to 86%; 95% CI = (66.4%, 75.8%)]. The prediction market beliefs and survey beliefs are highly correlated (Spearman correlation coefficient = 0.79, *P* < 0.001, *n* = 18). Both the prediction market beliefs (Spearman correlation coefficient = 0.30, *P* = 0.232, *n* = 18) and the survey beliefs (Spearman correlation coefficient 0.52, *P* = 0.028, *n* = 18) are positively correlated with the ranked degree of replication success.
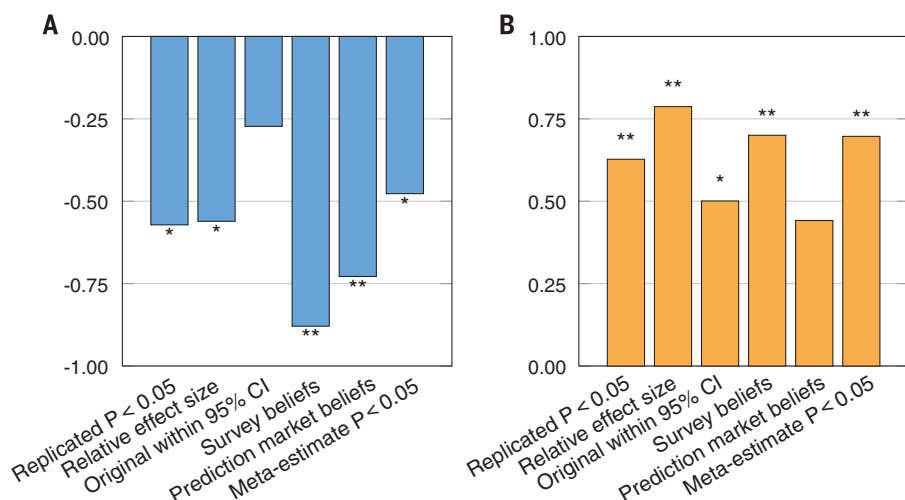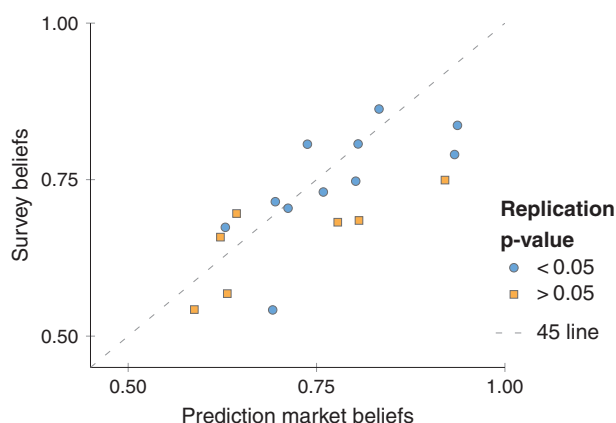
Fig. 3. Correlations between *P* values and sample sizes in original studies and replicability indicators. (**A**) The original *P* value is negatively correlated with all six replicability indicators, and five of these correlations are significant. (**B**) The original sample size is positively correlated with all six replicability indicators, and five of these correlations are significant. Spearman correlation coefficients are shown on the vertical axes. **P* < 0.05; **P < 0.01.
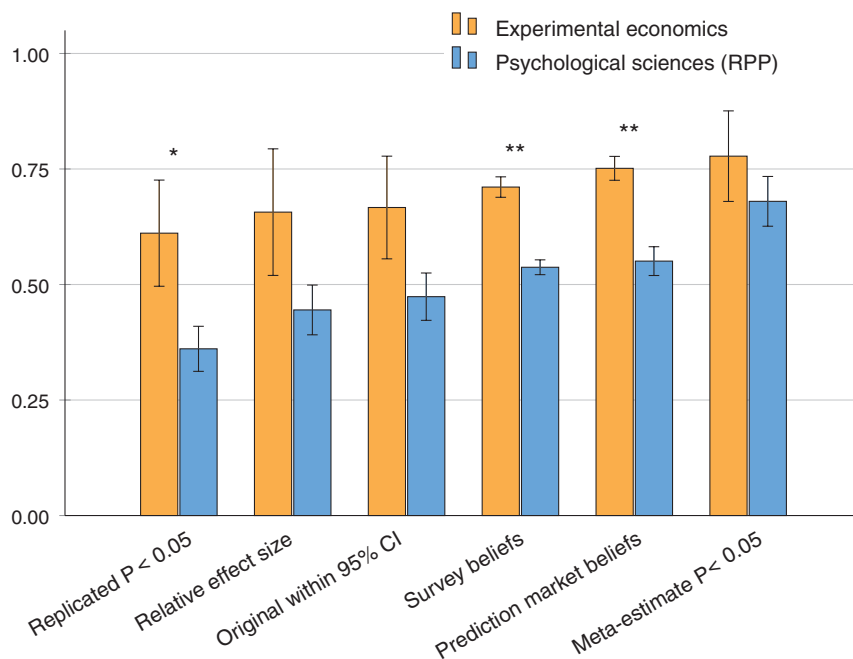
Fig. 4. A comparison of replicability indicators in experimental economics (this study) and psychological sciences (RPP). The graph shows means ± SE for replicability indicators. All six replicability indicators are higher for experimental economics; this difference is significant for three of the replicability indicators. The average difference in replicability across the six indicators is 19 percentage points. Details about the statistical tests are included in the supplementary materials. $*P < 0.05$; $**P < 0.01$.

thereby improving replicability. Second, pioneering experimental economists were eager for others to adopt their methods; to this end, they persuaded journals to print instructions and even original data. These editorial practices created norms of transparency and have made replication and reanalysis relatively easy.

There is every reason to be optimistic that science in general, and social science in particular, will emerge much improved after the current period of critical self-reflection. Our study suggests that laboratory experiments published in top economic journals have relatively high rates of replicability. Challenges still remain: For example, executing replications can be laborious, even when scientific journals require online posting of data and computer code to make things easier. This is a reminder that as scientists, we should design and document our methods to anticipate replication and make it easy to do. Our results also show that there is some information in post-publication peer beliefs (revealed in both markets and surveys), and perhaps even more information in simple statistics from published results, about whether studies are likely to be replicable. All of these developments suggest that the cultivation of good professional norms, discouragement of bad norms, policing of disclosure requirements by journals, and simple evidence-based editorial policies can improve scientific replicability, perhaps very quickly.

REFERENCES AND NOTES

1. M. McNutt, *Science* **343**, 229 (2014).
2. J. P. A. Ioannidis, *PLOS Med.* **2**, e124 (2005).
3. F. Prinz, T. Schlange, K. Asadullah, *Nat. Rev. Drug Discov.* **10**, 712 (2011).
4. C. G. Begley, L. M. Ellis, *Nature* **483**, 531–533 (2012).
5. L. P. Freedman, I. M. Cockburn, T. S. Simcoe, *PLOS Biol.* **13**, e1002165 (2015).
6. K. S. Button *et al.*, *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
7. J. K. Hewitt, *Behav. Genet.* **42**, 1–2 (2012).
8. M. S. Lawrence *et al.*, *Nature* **499**, 214–218 (2013).
9. E. E. Leamer, *Am. Econ. Rev.* **73**, 31 (1983).
10. A. E. Roth, *Empir. Econ.* **19**, 279–289 (1994).
11. Z. Maniadis, F. Tufano, J. A. List, *Am. Econ. Rev.* **104**, 277–290 (2014).
12. J. P. Simmons, L. D. Nelson, U. Simonsohn, *Psychol. Sci.* **22**, 1359–1366 (2011).
13. S. Carpenter, *Science* **335**, 1558–1561 (2012).
14. Open Science Collaboration, *Perspect. Psychol. Sci.* **7**, 657–660 (2012).
15. J. Bohannon, *Science* **344**, 788–789 (2014).
16. S. Doyen, O. Klein, C.-L. Pichon, A. Cleeremans, *PLOS ONE* **7**, e29081 (2012).
17. S. J. Ritchie, R. Wiseman, C. C. French, *PLOS ONE* **7**, e33423 (2012).
18. B. A. Nosek *et al.*, *Science* **348**, 1422–1425 (2015).
19. Open Science Collaboration, *Science* **349**, aac4716 (2015).
20. A. Gelman, H. Stern, *Am. Stat.* **60**, 328–331 (2006).
21. G. Cumming, *Perspect. Psychol. Sci.* **3**, 286–300 (2008).
22. R. Verhagen, E.-J. Wagenmakers, *J. Exp. Psychol. Gen.* **143**, 1457–1475 (2014).
23. U. Simonsohn, *Psychol. Sci.* **26**, 559–569 (2015).
24. J. T. Leek, P. Patil, R. D. Peng, http://arxiv.org/abs/1509.08968 (2015).
25. K. J. Arrow *et al.*, *Science* **320**, 877–878 (2008).
26. A. Dreber *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15343–15347 (2015).
27. J. Wolfers, E. Zitzewitz, *Interpreting Prediction Market Prices as Probabilities* (Working Paper No. 12200, National Bureau of Economic Research, 2006).
28. J. P. A. Ioannidis, *Epidemiology* **19**, 640–648 (2008).
29. B. D. McCullough, H. D. Vinod, *Am. Econ. Rev.* **93**, 873–892 (2003).
30. W. G. Dewald, J. G. Thursby, R. G. Anderson, *Am. Econ. Rev.* **76**, 587 (1986).
31. B. D. McCullough, K. A. McGeary, T. D. Harrison, *J. Money Credit Bank.* **38**, 1093–1107 (2006).
32. A. Brodeur, M. Lé, M. Sangnier, Y. Zylberberg, *AEJ Applied* **8**, 1–32 (2016).
33. J. Abeler, A. Falk, L. Goette, D. Huffman, *Am. Econ. Rev.* **101**, 470–492 (2011).
34. A. Ambrus, B. Greiner, *Am. Econ. Rev.* **102**, 3317–3332 (2012).
35. B. Bartling, E. Fehr, K. M. Schmidt, *Am. Econ. Rev.* **102**, 834–864 (2012).
36. G. Charness, M. Dufwenberg, *Am. Econ. Rev.* **101**, 1211–1237 (2011).
37. R. Chen, Y. Chen, *Am. Econ. Rev.* **101**, 2562–2589 (2011).
38. G. de Clippel, K. Eliaz, B. G. Knight, *Am. Econ. Rev.* **104**, 3434–3458 (2014).
39. J. Duffy, D. Puzzello, *Am. Econ. Rev.* **104**, 1735–1776 (2014).
40. U. Dulleck, R. Kerschbamer, M. Sutter, *Am. Econ. Rev.* **101**, 526–555 (2011).
41. K. M. Marzilli Ericson, A. Fuster, *Q. J. Econ.* **126**, 1879–1907 (2011).
42. E. Fehr, H. Herz, T. Wilkening, *Am. Econ. Rev.* **103**, 1325–1359 (2013).
43. D. Friedman, R. Oprea, *Am. Econ. Rev.* **102**, 337–363 (2012).
44. D. Fudenberg, D. G. Rand, A. Dreber, *Am. Econ. Rev.* **102**, 720–749 (2012).
45. S. Huck, A. J. Seltzer, B. Wallace, *Am. Econ. Rev.* **101**, 819–843 (2011).
46. J. Ifcher, H. Zarghamee, *Am. Econ. Rev.* **101**, 3109–3129 (2011).
47. J. B. Kessler, A. E. Roth, *Am. Econ. Rev.* **102**, 2018–2047 (2012).
48. M. Kirchler, J. Huber, T. Stöckl, *Am. Econ. Rev.* **102**, 865–883 (2012).
49. S. Kogan, A. M. Kwasnica, R. A. Weber, *Am. Econ. Rev.* **101**, 927–947 (2011).
50. I. Kuziemko, R. W. Buell, T. Reich, M. I. Norton, *Q. J. Econ.* **129**, 105–149 (2014).

Science

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen and Hang Wu

Another social science looks at itself
Experimental economists have joined the reproducibility discussion by replicating selected published experiments from two top-tier journals in economics. Camerer *et al.* found that two-thirds of the 18 studies examined yielded replicable estimates of effect size and direction. This proportion is somewhat lower than unaffiliated experts were willing to bet in an associated prediction market, but roughly in line with expectations from sample sizes and P values.
Science, this issue p. 1433

ARTICLE TOOLS	http://science.sciencemag.org/content/351/6280/1433
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2016/03/02/science.aaf0918.DC1
REFERENCES	This article cites 62 articles, 8 of which you can access for free http://science.sciencemag.org/content/351/6280/1433#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service