

Large teams develop and small teams disrupt science and technology

Lingfei Wu^{1,2}, Dashun Wang^{3,4,5} & James A. Evans^{1,2,6*}

One of the most universal trends in science and technology today is the growth of large teams in all areas, as solitary researchers and small teams diminish in prevalence^{1–3}. Increases in team size have been attributed to the specialization of scientific activities³, improvements in communication technology^{4,5}, or the complexity of modern problems that require interdisciplinary solutions^{6–8}. This shift in team size raises the question of whether and how the character of the science and technology produced by large teams differs from that of small teams. Here we analyse more than 65 million papers, patents and software products that span the period 1954–2014, and demonstrate that across this period smaller teams have tended to disrupt science and technology with new ideas and opportunities, whereas larger teams have tended to develop existing ones. Work from larger teams builds on more-recent and popular developments, and attention to their work comes immediately. By contrast, contributions by smaller teams search more deeply into the past, are viewed as disruptive to science and technology and succeed further into the future—if at all. Observed differences between small and large teams are magnified for higher-impact work, with small teams known for disruptive work and large teams for developing work. Differences in topic and research design account for a small part of the relationship between team size and disruption; most of the effect occurs at the level of the individual, as people move between smaller and larger teams. These results demonstrate that both small and large teams are essential to a flourishing ecology of science and technology, and suggest that, to achieve this, science policies should aim to support a diversity of team sizes.

Advocates of team science have claimed that a shift to larger teams in science and technology fulfils the essential function of solving problems in modern society that are complex and which require interdisciplinary solutions^{6–8}. Although much has been demonstrated about the professional and career benefits of team size for team members⁹, there is little evidence that supports the notion that larger teams are optimized for knowledge discovery and technological invention⁹. Experimental and observational research on groups reveals that individuals in large groups think and act differently—they generate fewer ideas^{10,11}, recall less learned information¹², reject external perspectives more often¹³ and tend to neutralize each other's viewpoints¹⁴. Small and large teams may also differ in their response to the risks associated with innovation. Large teams, such as large business organizations, may focus on sure bets with large potential markets, whereas small teams that have more to gain and less to lose may undertake new, untested opportunities with the potential for high growth and failure¹⁵, leading to markedly different outcomes. These possibilities led us to explore the consequences of smaller and larger teams for scientific and technological advance, and how such teams search and assemble knowledge differently.

Previous research demonstrates that large article and patent teams receive slightly more citations^{2,16}. However, citation counts alone cannot capture distinct types of contribution. This can be seen in the

difference between two well-known articles: one about self-organized criticality¹⁷ (the BTW model, after the authors' initials) and another about Bose–Einstein condensation¹⁸ (for which Wolfgang Ketterle was awarded the 2001 Nobel Prize in Physics) (Fig. 1, Extended Data Fig. 1b). The two articles have received a similar number of citations, but most research subsequent to the BTW-model article has cited only the model itself without mentioning references from the article. By contrast, the Bose–Einstein condensation article is almost always co-cited with Bose¹⁹, Einstein²⁰ and other antecedents. The difference between the two papers is reflected not in citation counts but in whether they suggested or solved scientific problems—whether they disrupted or developed existing scientific ideas, respectively²¹. The BTW model launched new streams of research, whereas the experimental realization of Bose–Einstein condensation elaborated upon possibilities that had previously been posed.

To systematically evaluate the role that small and large teams have in unfolding scientific and technological advances, we collected large-scale datasets from three related but distinct domains (see Methods): (1) the Web of Science (WOS) database that contains more than 42 million articles published between 1954 and 2014, and 611 million citations among them; (2) 5 million patents granted by the US Patent and Trademark Office from 1976 to 2014, and 65 million citations added by patent applicants; (3) 16 million software projects and 9 million forks to them on GitHub (2011–2014), a popular web platform that allows users to collaborate on the same code repository and 'cite' other repositories by copying and building on their code.

For each dataset, we assess the degree to which each work disrupts the field of science or technology to which it belongs by introducing something new that eclipses attention to previous work upon which it has built. We use a measure that was previously designed²² to identify destabilization and consolidation in patented inventions; this measure varies between -1 and 1 , which corresponds to science and technology that develops or disrupts, respectively (Fig. 1a). We validate the disruption measure in several ways. First, we investigate the distribution of disruption across scientific papers (Fig. 1b); the disruptive BTW-model article is located in the top 1%, whereas the developmental Bose–Einstein condensation paper is in the bottom 3% of the disruption distribution. We also find that, on average, Nobel-prize-winning papers register among the 2% most disruptive articles. Review articles are developmental with a negative mean of disruption (bottom 46%), whereas the original research works that they review have a positive mean (top 23%). Articles that headline prominent prior work—such as the Bose–Einstein condensation article—lie in the bottom 25% (Supplementary Table 1). We further confirmed these results with a survey in which we asked scholars from diverse fields to propose disruptive and developmental articles; this symmetrically confirmed the disruption measure (Supplementary Table 2). Finally, we find that in the titles of articles different words associate with disruptive ('introduce', 'measure', 'change' and 'advance') versus developing ('endorse', 'confirm', 'demonstrate', 'theory' and 'model') papers (Fig. 1c, Supplementary Table 3).

¹Department of Sociology, University of Chicago, Chicago, IL, USA. ²Knowledge Lab, University of Chicago, Chicago, IL, USA. ³Kellogg School of Management, Northwestern University, Evanston, IL, USA. ⁴Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA. ⁵McCormick School of Engineering, Northwestern University, Evanston, IL, USA. ⁶Santa Fe Institute, Santa Fe, NM, USA. *e-mail: jevans@uchicago.edu

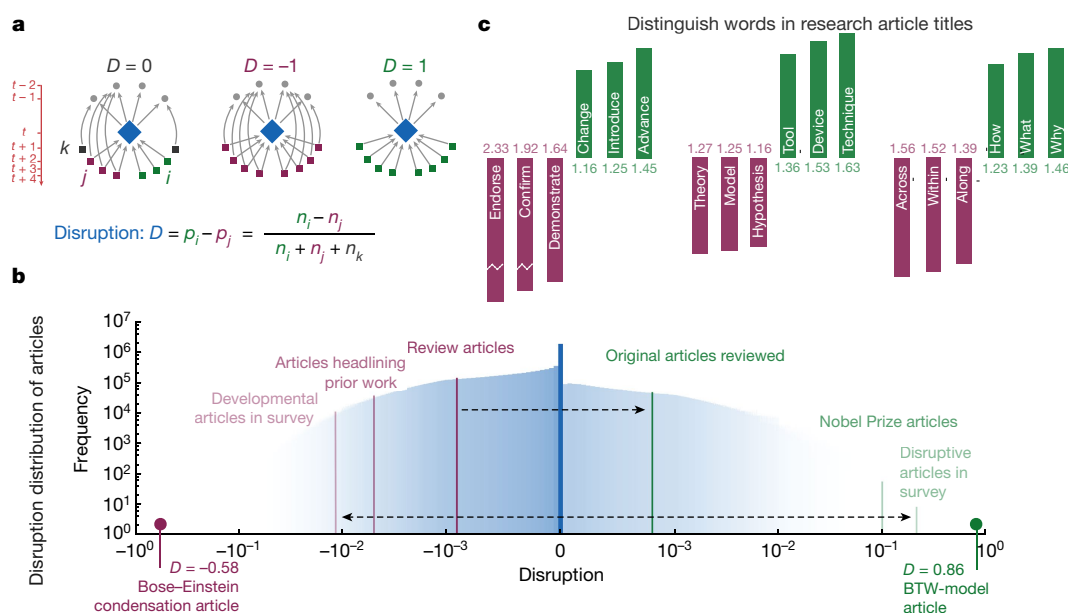


Fig. 1 | Quantifying disruption. **a**, Simplified illustration of disruption. Three citation networks comprising focal papers (blue diamonds), references (grey circles) and subsequent work (rectangles). Subsequent work may cite the focal work (i , green), both the focal work and its references (j , red) or just its references (k , black). Disruption, D , of the focal paper is defined by the difference between the proportion of type i and j papers $p_i - p_j$, which equals the difference between the observed number of these papers $n_i - n_j$ divided by the number of all subsequent works $n_i + n_j + n_k$. A paper may be disrupting ($D = 1$), neutral ($D = 0$) or developing ($D = -1$). **b**, The distribution of disruption across 25,988,101 WOS journal articles published between 1900 and 2014. On this distribution, we mark the BTW-model ($D = 0.86$, top 1%) and Bose-Einstein condensation articles ($D = -0.58$, bottom 3%) along with several samples used to validate D (Methods, Supplementary Tables 1–3). This includes (1) 104 ‘disruptive’ articles (disruption mean $E(D) = 0.215$,

top 2%) and 86 ‘developing’ articles ($E(D) = -0.011$, bottom 13%) nominated by a surveyed panel of 20 scholars across fields; (2) 877 Nobel-prize-winning papers published between 1902 and 2009 ($E(D) = 0.10$, top 2%); (3) 22,672 review articles ($E(D) = -0.0009$, bottom 46%) and 1,338,808 original research articles that they review ($E(D) = 0.0008$, top 23%); and (4) 148,303 articles that headline prominent prior work by mentioning one or more cited authors in the title ($E(D) = -0.0049$, bottom 24%). **c**, We select titles from 24,174,022 articles published between 1954 and 2014 and assign them to one of two groups, disrupting ($D > 0$) or developing ($D < 0$) articles. For the 1,033,879 words observed in both groups, we calculate the ratio of frequency in disrupting versus developing articles, r . We visualize differences in the content and writing style between these two groups in terms of verbs, nouns, and adverbs and prepositions (from left to right). To facilitate comparison, we visualize r in green if $r > 1$, and $1/r$ in red otherwise.

We predict that work by small teams will be substantially more disruptive than work by large teams. Our databases of papers, patents and software strongly confirm this prediction. Our sources differ in scope and domain, but we consistently observe that over the past 60 years, larger teams produce articles, patents and software with a disruption score that markedly and monotonically declines with each additional team member (Fig. 2a–c, Extended Data Fig. 3). Specifically, as teams grow from 1 to 50 team members, their papers, patents and products drop in percentiles of measured disruption by 70, 30 and 50, respectively (Extended Data Fig. 3a). In every case, this highlights a transition from disruption to development. These results support the hypothesis that large teams may be better designed or incentivized to develop current science and technology, and that small teams disrupt science and technology with new problems and opportunities.

This phenomenon is amplified when we focus on the most disruptive and impactful work (Fig. 2d–f). We measure the impact of each article, patent and software using the number of citations each work received. As shown in Fig. 2d, solo authors are just as likely to produce high-impact papers (in the top 5% of citations) as teams with five members, but solo-authored papers are 72% more likely to be highly disruptive (in the top 5% of disruptive papers). By contrast, ten-person teams are 50% more likely to score a high-impact paper, yet these contributions are much more likely to develop existing ideas already prominent in the system, which is reflected in the very low likelihood they are among the most disruptive. By repeating the same analyses for patents (Fig. 2e) and software development (Fig. 2f), we find that disruption and impact consistently diverge as teams grow in size.

Differences in disruption between works produced by small and large teams are magnified as the impact of the work increases (Fig. 3a); high-impact papers produced by small teams are the most disruptive,

and high-impact papers produced by large teams are the most developmental. As article impact increases, the negative slope of disruption as a function of team size steepens sharply. Even within the pool of high-impact articles and patents (Fig. 3a, top 5% of citations), which are statistically more likely produced by large teams (Fig. 2d), small teams have disrupted the current system with substantially more new ideas. We further split papers by time period (Extended Data Fig. 3c) and scientific field (Fig. 3b, Extended Data Fig. 4), and found that these patterns linking disruption and team size are stable for all eras and for 90% of disciplines. The only consistent exceptions were observed for engineering and computer science, in which conference proceedings rather than journal articles are the publishing norm (the WOS database indexes only journal articles).

We considered whether observed differences between the work of small and large teams could simply be attributed to differences in disruptive potential for the different types of articles that they produce; for example, small teams may generate more theoretical innovations and large teams more empirical analyses. Drawing on a previous approach²³, we matched papers from www.arXiv.org with the WOS database and repeated our analyses controlling for the number of figures in each article (Extended Data Fig. 5a), as empirical papers tend to have more figures than theoretical ones. Our results suggest that most of the difference in disruption between work from smaller and larger teams is not driven by differences in whether they contributed theoretical versus empirical papers (that is, had more or less figures). The association remains the same when we consider other distinctions, including review versus original research articles. Review articles with fewer authors are more disruptive than those with more, just as with original research articles (Extended Data Fig. 5b).

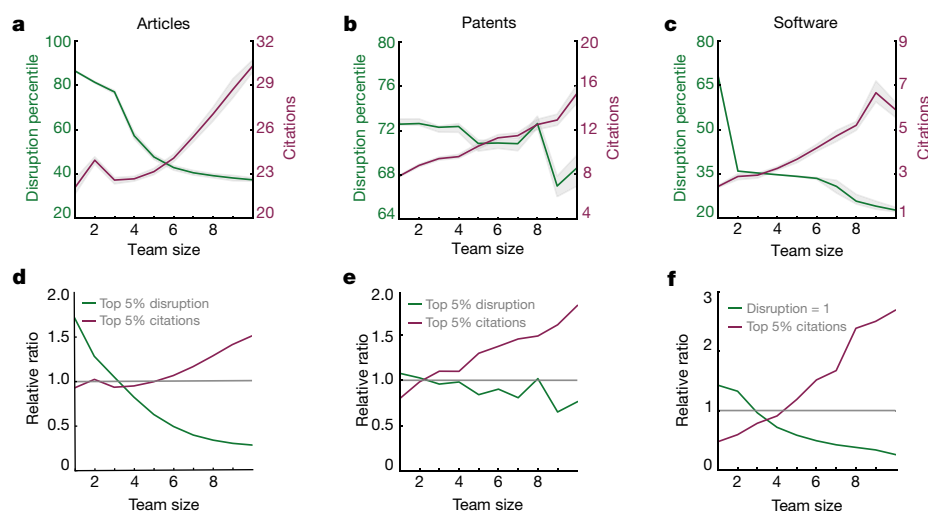


Fig. 2 | Small teams disrupt whereas large teams develop. **a–c**, For research articles (24,174,022 WOS articles published between 1954 and 2014), patents (2,548,038 US patents assigned between 2002 and 2014) and software (26,900 GitHub repositories uploaded between 2011 and 2014), median citations (red curves, indexed by right y axis) increase with team size whereas the average disruption percentile (green curves, indexed by left y axis) decreases with team size. For all datasets, we present work with one or more citations. Teams of between 1 and 10 authors account for 98% of articles, 99% of patents and 99% of code repositories. Bootstrapped 95% confidence intervals are shown as grey zones. Extended Data Figure 3a shows that observed relationships hold for two orders of magnitude of team size. **d–f**, As in **a–c** but for extreme cases rather than for average behaviour. Relative ratios compare the observed proportion of teamwork being extremely (top 5%) disruptive or impactful (measured with

citations) against a constant baseline (grey line $y = 1$), which indicates a situation in which the most disruptive and impactful work is distributed equally across team sizes. We find that the probability of observing papers, patents and products of highest impact increases with team size (Kolmogorov–Smirnov statistics and probabilities for all team sizes plotted in Extended Data Fig. 2f), whereas the probability of observing the most disruptive work decreases with team size (t -statistics and probabilities for all team sizes plotted in Extended Data Fig. 2c). For example, **d** shows that the percentage of top 5% disruptive papers depends on team size, with 8.6% contributed by single authors and only 1.4% contributed by teams of ten authors. This posts relative ratios of $8.6/5 = 1.72$ and $1.4/5 = 0.28$, respectively. For software, 69% of the codebases have disruption values that equal 1; we therefore use this maximum value instead of the top 5%.

Another possible explanation for our results is that the team effect that we observe occurs because the scientists, inventors and software designers involved in larger teams are qualitatively different from those comprising smaller teams. But when we predict disruptiveness as a function of team size, controlling for publication year, topic and author (Fig. 3c, Extended Data Fig. 3b, Supplementary Table 4), we find that the decrease of disruption with the growth of team size continues to hold, and controlling for authors greatly improves the percentage of variance explained (Supplementary Table 4).

We further test the robustness of our results against several different definitions of the disruption measure, including the removal of

self-citation links, exclusion of all but high-impact references and other variations (Extended Data Fig. 5g–i). Across all variations, our conclusions remain the same.

The considerable difference in disruption between large and small teams raises questions regarding how these teams differ in searching the past to formulate their next paper, patent or product. When we dissect search behaviour, we find that large and small teams engage in notably different practices that may be related to divergent contributions in disruption and impact. Specifically, we measure search depth as average relative age of references cited and search popularity as median citations to the references of a focal work. We examine these search strategies

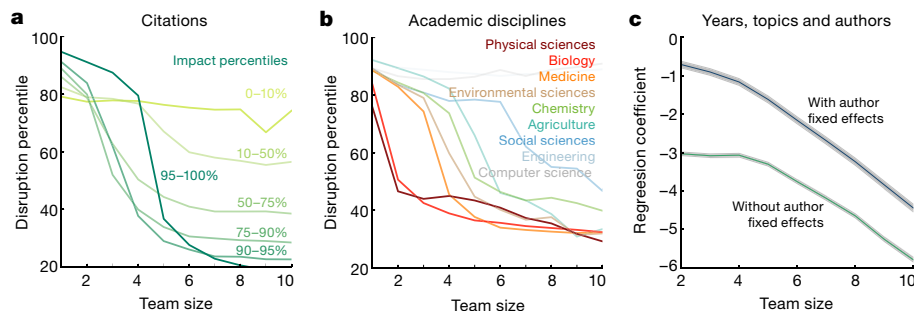


Fig. 3 | Small teams disrupt across impact levels and fields. **a**, The disruption percentile decreases with team size across impact levels for research articles (24,174,022 WOS articles published between 1954 and 2014). Curves are coloured by impact percentile (in number of citations). The disruption percentile decreases faster for higher-impact articles (darker green curves). The transition from disruption to development ($D = 0$) occurs when the disruption percentile equals 70. **b**, Disruption decreases with team size across nine fields for research articles. These fields were manually coded, on the basis of 258 sub-field labels attached to journals in WOS data. **c**, Plot of the regression coefficients of disruption

percentile on team size from linear regressions, controlling for publication year, topic and author. The regression is based on the 96,386,516 WOS research articles (articles are counted repeatedly if they appear across the publication records of different scholars), contributed by 38,000,470 name-disambiguated scholars. To control for topics, we use the field codes inherited from **b**. Estimated parameters from the regression models are presented in Supplementary Table 4. The same regressions using raw values of disruption rather than disruption percentile are shown in Extended Data Fig. 3b.

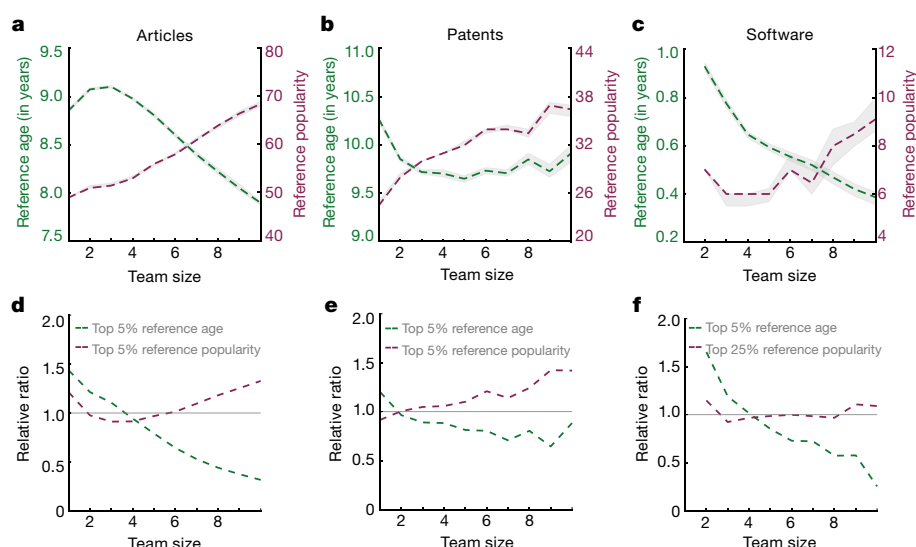


Fig. 4 | Small and large teams behave differently in their search through past work. **a–c**, For research articles (24,174,022 WOS articles published between 1954 and 2014), patents (2,548,038 US patents granted between 2002 and 2014) and software (26,900 GitHub repositories uploaded between 2011 and 2014), the median popularity of references (in number of citations, shown as red curves and indexed by the right y axis) increases with team size, whereas the average age of references (green curves, indexed by the left y axis) decreases with team size. For all datasets, we present work with one or more citations. Bootstrapped 95% confidence intervals are shown as grey zones. Teams of between 1 and 10 authors account for 98% of articles, 99% of patents and 99% of code repositories. Extended Data Figure 3a shows that the observed relationships hold for two orders of magnitude of team size. **d–f**, As in **a–c**, but for extreme cases rather than for average behaviour. Relative ratios compare empirically

observed proportions of teamwork that searches for extremely early or unpopular previous ideas against theoretical baselines of what would have been expected at random. The grey line ($y = 1$) indicates a scenario in which work building upon the earliest and the most unpopular ideas is distributed equally across team sizes. We find that the probability of observing papers, patents and products built upon the most influential previous work increases with team size, whereas the probability of citing older work decreases with team size. For example, **d** shows that the percentage of the 5% of articles that cite the oldest ideas is unequally distributed, with 7.2% contributed by single authors and only 1.6% contributed by ten author teams. This provides relative ratios $7.2/5 = 1.44$ and $1.6/5 = 0.32$, respectively. Software has very few high-citation codebases; we therefore use the top 25% rather than top 5% reference popularity for our calculations.

across fields, time periods and impact levels in science, technology and software. We also relate these search strategies to temporal delay in the impact these works receive using the ‘Sleeping Beauty index’²⁴, which captures a delayed burst of attention traced by convexity in the citation attention that a work receives over time.

We find that solo authors and small teams much more often build on older, less popular ideas (Fig. 4, Extended Data Fig. 6). Larger teams more often target recent, high-impact work as their primary source of inspiration, and this tendency increases monotonically with team size. It follows that large teams receive more of their citations rapidly, as their work is immediately relevant to more contemporaries whose ideas they develop and audiences primed to appreciate them. Conversely, smaller teams experience a much longer citation delay; the average Sleeping Beauty index percentile for solo and two-person research teams is twice that of ten-person teams (Extended Data Fig. 7). As a result, even though small teams receive less recognition overall owing to the rapid decay of collective attention^{25–27} (as shown in Fig. 2a), their successful research produces a ripple effect, which becomes an influential source of later large-team success (Extended Data Fig. 8).

We also consider the relationship between these distinctive search mechanisms and recent findings²⁸ that suggest multi- and interdisciplinary teams more often link work from divergent fields. We examined the novelty of journal combinations within article reference lists and also keyword combinations within articles in relation to team size. These show consistent diminishing marginal increases to novelty with team size, such that with each new team member, their contribution to novel combinations decreases (Extended Data Fig. 9). Moreover, using a previous measure of atypical combinations²⁸, we find that atypical combinations increase slowly up to teams of approximately ten and then decrease sharply below the value associated with a solo investigator. Whereas larger teams facilitate broader search, small teams search deeper.

In summary, we report a universal and previously undocumented pattern that systematically differentiates the contributions of small and large teams in the creation of scientific papers, technology patents and software products. Small teams disrupt science and technology by exploring and amplifying promising ideas from older and less-popular work. Large teams develop recent successes, by solving acknowledged problems and refining common designs. Some of this difference results from the substance of science and technology that small versus large teams tackle, but the larger part appears to emerge as a consequence of team size itself. Certain types of research require the resources of large teams, but large teams demand an ongoing stream of funding and success to ‘pay the bills’²⁹, which makes them more sensitive to the loss of reputation and support that comes from failure³⁰. Our findings are consistent with field research on teams in other domains, which demonstrate that small groups with more to gain and less to lose are more likely to undertake new and untested opportunities that have the potential for high growth and failure¹⁵. Our findings are also in accordance with experimental and observational research on groups that demonstrates how individuals in large groups think and act differently from those in small groups^{10–14}.

Both small and large teams are essential to a flourishing ecology of science and technology. Taken together, the increasing dominance of large teams, a flurry of scholarship on their perceived benefits^{2,6–9,28,31} and our findings call for new investigations into the vital role that individuals and small groups have in advancing science and technology. Direct sponsorship of small-group research may not be enough to preserve its benefits. We analysed articles published from 2004 to 2014 that acknowledged financial support from several top government agencies around the world, and found that small teams with this funding are indistinguishable from large teams in their tendency to develop rather than disrupt their fields (Extended Data Fig. 10). In contrast to Nobel Prize papers, which have an average disruption among the top 2% of all contemporary papers, funded papers rank near the bottom 31%. This

could result from a conservative review process, proposals designed to anticipate such a process or a planning effect whereby small teams lock themselves into large-team inertia by remaining accountable to a funded proposal. When we compare two major policy incentives for science (funding versus awards), we find that Nobel-prize-winning articles significantly oversample small disruptive teams, whereas those that acknowledge US National Science Foundation funding oversample large developmental teams. Regardless of the dominant driver, these results paint a unified portrait of underfunded solo investigators and small teams who disrupt science and technology by generating new directions on the basis of deeper and wider information search. These results suggest the need for government, industry and non-profit funders of science and technology to investigate the critical role that small teams appear to have in expanding the frontiers of knowledge, even as large teams rapidly develop them.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-0941-9>.

Received: 20 April 2018; Accepted: 7 January 2019;

Published online 13 February 2019.

- Guimerà, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).
- Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
- Hunter, L. & Leahey, E. Collaborative research in sociology: trends and contributing factors. *Am. Sociol.* **39**, 290–306 (2008).
- Jones, B. F., Wuchty, S. & Uzzi, B. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).
- Xie, Y. “Undemocracy”: inequalities in science. *Science* **344**, 809–810 (2014).
- Milojević, S. Principles of scientific research team formation and evolution. *Proc. Natl Acad. Sci. USA* **111**, 3984–3989 (2014).
- Falk-Krzesinski, H. J. et al. Mapping a research agenda for the science of team science. *Res. Eval.* **20**, 145–158 (2011).
- Committee on the Science of Team Science. *Enhancing the Effectiveness of Team Science* (National Academies Press, Washington DC, 2015).
- Leahey, E. From sole investigator to team scientist: trends in the practice and study of research collaboration. *Annu. Rev. Sociol.* **42**, 81–100 (2016).
- Paulus, P. B., Kohn, N. W., Arditti, L. E. & Korde, R. M. Understanding the group size effect in electronic brainstorming. *Small Group Res.* **44**, 332–352 (2013).
- Lakhani, K. R. et al. Prize-based contests can provide solutions to computational biology problems. *Nat. Biotechnol.* **31**, 108–111 (2013).
- Barber, S. J., Harris, C. B. & Rajaram, S. Why two heads apart are better than two heads together: multiple mechanisms underlie the collaborative inhibition effect in memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **41**, 559–566 (2015).
- Minson, J. A. & Mueller, J. S. The cost of collaboration: why joint decision making exacerbates rejection of outside information. *Psychol. Sci.* **23**, 219–224 (2012).
- Greenstein, S. & Zhu, F. Open content, Linus’ law, and neutral point of view. *Inf. Syst. Res.* **27**, 618–635 (2016).
- Christensen, C. M. *The Innovator’s Dilemma: The Revolutionary Book That Will Change the Way You Do Business* (Harper Business, New York, 2011).
- Klug, M. & Bagrow, J. P. Understanding the group dynamics and success of teams. *R. Soc. Open Sci.* **3**, 160007 (2016).
- Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: an explanation of the $1/f$ noise. *Phys. Rev. Lett.* **59**, 381–384 (1987).
- Davis, K. B. et al. Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
- Bose, S. N. Plancks Gesetz und Lichtquantenhypothese. *Z. Physik* **26**, 178–181 (1924).
- Einstein, A. Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* **1**, 3 (1925).
- March, J. G. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**, 71–87 (1991).
- Funk, R. J. & Owen-Smith, J. A dynamic network measure of technological change. *Manage. Sci.* **63**, 791–817 (2017).
- Moody, J. The structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999. *Am. Sociol. Rev.* **69**, 213–238 (2004).
- Ke, Q., Ferrara, E., Radicchi, F. & Flammini, A. Defining and identifying Sleeping Beauties in science. *Proc. Natl Acad. Sci. USA* **112**, 7426–7431 (2015).
- Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
- Evans, J. A. Electronic publication and the narrowing of science and scholarship. *Science* **321**, 395–399 (2008).
- Gerow, A., Hu, Y., Boyd-Graber, J., Blei, D. M. & Evans, J. A. Measuring discursive influence across scholarship. *Proc. Natl Acad. Sci. USA* **115**, 3308–3313 (2018).
- Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
- Kuhn, T. S. The function of measurement in modern physical science. *Isis* **52**, 161–193 (1961).
- Collins, D. *Organizational Change: Sociological Perspectives* (Routledge, New York, 1998).
- Jones, B. F. The burden of knowledge and the ‘death of the Renaissance man’: is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).

Acknowledgements We are grateful for support from AFOSR grants FA9550-15-1-0162 and FA9550-17-1-0089, the John Templeton Foundation’s grant to the Metaknowledge Network, DARPA’s Big Mechanism program grant 14145043, National Science Foundation grant SBE 1158803, 1829344 and 1829366. We thank the University of Chicago Organizations and Markets seminar, the Swarna Club (Beijing), and Clarivate Analytics for supplying the Web of Science data.

Reviewer information Nature thanks L. Bornmann, S. Wuchty and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions L.W., D.W. and J.A.E. collaboratively conceived and designed the study, contributed data for analysis, and drafted, revised and edited the manuscript. L.W. analysed the data, ran all models and produced all visualizations.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-019-0941-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-0941-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.A.E.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

No statistical methods were used to predetermine sample size. Randomization and blinding were not possible, given the observational nature of the study.

Dataset of research articles. Our WOS dataset contains 43,661,387 journal papers and 615,697,434 citations that span from 1900 to 2014. These papers are published across 15,146 journals. Data before the 1950s are sparse, and so results presented in the main text focus on papers published between 1954 and 2014. Data from this period contain 42,045,077 papers distributed across 15,070 journals. Among these articles, 66% (27,728,266) are cited at least once, generating 611,483,153 citations in total. To calculate disruption and other network measures, we constructed a directed network with papers as nodes and citations as links. We calculated the disruption score for 25,988,101 papers published between 1900 and 2014, among which are 24,174,022 papers published between 1954 and 2014.

Dataset of patents. The US Patent and Trademark Office patent dataset contains 4,910,816 patents and 64,694,807 citations between 1976 and 2014, which represents the portion of the dataset with curated digital patent application data. Citation links added by inventors and patent examiners represent different dynamics; examiner citations do not reflect the technology on which a proposed invention is built but rather the technologies with which it competes³². As such, we focus only on applicant citations, which are marked in the dataset after 2001 and represent 53% of total citations. From 2002 to 2014, we have 2,548,038 patents in total that are linked by 44,798,680 inventor citations. To calculate disruption and other measures, we constructed a directed network that contained patents as nodes and applicant citations as links.

Dataset of software. The GitHub data contain 15,984,275 code bases (or repositories) contributed by 2,348,085 programmers in GitHub between 2011 and 2014. In this period, 2,065,729 programmers contributed 9,127,410 forking patterns in which they copied and saved an existing repository to build upon it. To calculate disruption and other measures, we construct a citation network of repositories. For each repository, we identify its core members as those who contributed more edits, or 'pushes', than the average value of all contributors to a repository¹⁶. We then add a citation link from repository A to B if a core member of A forked the code from B between this user's first and last edit of A. The constructed network contains 26,900 nodes (repositories) and 108,640 links.

Dataset of name-disambiguated WOS scholars. We use a hybrid algorithm to exploit both metadata and citations in disambiguating WOS authors. For each name (including family name and initials), we construct a network of relevant papers connected on the basis of a similarity measure that considers shared co-authors, references and citations³³. Disconnected components of this network are assumed to correspond to distinct authors. As co-citation is an important feature in this similarity measure, our algorithm applies only to the 28,607,001 cited papers in the whole dataset of 43,661,387 papers (1900–2014). Different from a previous study³³, we also use emails and institutions of authors to improve the algorithm by connecting name clusters that share this information. Although only 3% of the cited papers have email information and/or organization information, these papers connect 72% of the remaining papers. As emails are unique and institutions are rarely shared by scholars of the same name, adding metadata makes the unsupervised algorithm semi-supervised, reduces the time complexity and increases accuracy. Finally, we obtain 10,051,491 scholars who contributed to 22,177,224 papers. Eighty-five per cent of these scholars contributed to three or more papers, and 44% contributed to four or more. We use the 2017 Open Researcher and Contributor ID (ORCID) dataset to validate the name disambiguation results, and find that precision is 78% and recall is 86% among the 118,094 ORCID scholars with three or more papers. We also test the results using a dataset of 31,070 Chinese scholars and 253,786 papers retrieved from the project outcome reports of research funded by National Natural Science Foundation of China; precision found in this test is 79%, and recall is 65%.

Removing self-citations from WOS papers. Using the above-mentioned data of name-disambiguated scholars in WOS, we are able to test the robustness of the negative association between team size and disruption against the removal of self-citations (Extended Data Fig. 5). If a paper cites another that shares at least one common name-disambiguated author, we define it as a self-citation. Among the 615,697,434 citations created between 43,661,387 papers between 1900 and 2014, 10.2% (62,626,733) are self-citations. For the 28,607,001 papers with at least one citation, 36.3% of them benefit from one or more self-citations. The averaged percentage of self-citation increases monotonically with team size from 2.9% for single-author papers and 8.7% for two authors to 12.3% for three authors, and stabilizes at approximately 30% for 50 or more authors. This percentage also increases rapidly with the number of citations but peaks at 15% for ten-citation papers and then slowly decreases, returning to below 9% (which is the same percentage as in two-citation papers) for 100 or more citations.

Dataset of Nobel-prize-winning WOS papers. We collect 877 WOS papers, each of which earned their author(s) a Nobel Prize. These papers are published across 178 journals during the time period 1902 to 2009, including 316 papers in

Physiology or Medicine, 284 papers in Physics and 277 papers in Chemistry. The average disruption of the Nobel-prize-winning papers is 0.10, ranking in the top 2% of all WOS papers from the same time period.

Dataset of government-funded WOS papers. For the 43,661,387 WOS papers published between 1900 and 2014, WOS recorded acknowledged financial support for 10.9% (4,754,769). The percentage of financially supported papers began in 2008, following a commitment by the WOS to record this information, and accelerated from that time; 15.2% in 2008, 38.9% in 2009 and 55.8% in 2014³⁴. To analyse the disruption of government-funded papers, we select 477,702 WOS papers that acknowledged funding from five major government agencies, published between 2004 and 2014. The acknowledged agencies include the National Science Foundation (NSF; 191,717 papers), National European Research Council and European Commission (ERC and EC; 81,296 papers), Natural Science Foundation of China (NSFC; 80,448 papers), German Research Foundation (DFG; 75,881 papers), and Japan Society for the Promotion of Science (JSPS; 58,275 papers). These papers are published across 7,325 journals. A paper may be funded by multiple agencies. The average disruption of these papers is -0.0024 , ranking in the tail 31% of all WOS papers from the same period. For NSF-funded papers, we calculate the average grant size (over multiple NSF grants acknowledged by the same paper). We find 140,972 papers that were supported by grants smaller than 1 million US dollars, 24,370 papers that were supported by grants 1–5 million US dollars and 26,375 papers that were supported by grants of greater than 5 million US dollars.

Fields, subfields and journals of WOS papers. The articles that we analysed are published across 15,146 journals that belong to 258 subfields, according to the subject category labels for journals in the WOS dataset. We code these subfields into ten major fields that comprise the physical sciences, biology, medicine, environmental sciences, chemistry, agriculture, social sciences, engineering, computer science and other. In Fig. 3b, we show the average disruption percentile against team size across nine fields, except 'other'. In Extended Data Fig. 4, we selectively display the average disruption percentile against team size at the journal level for three or four subfields from each of the nine fields, except for computer science. We use ordinary least squares regression to fit the relation between team size and disruption percentile for 10,907 journals across 218 subfields. We find that among all studied journals, 86% post negative regression coefficients. If we only consider journals that publish a substantial number of articles or those for which the regression coefficient is significant, this fraction is higher: 91% of journals with more than 3,000 articles show a negative relationship between team size and disruption percentile, and 88% of journals post significant negative regression coefficients.

Modelling topics of WOS papers using Doc2vec. We randomly selected 100,000 papers from 15,146 journals, weighted by the frequency of articles published by those journals, to ensure that these papers cover a variety of topics. We then selected titles and abstracts from these papers and used them as the corpus on which to train a neural network that converts documents into vectors (Doc2vec)³⁵. We used the Gensim Python library to train the vector space with model parameters as follows: size = 100 (vector length), min_count = 2 (minimum frequency of words used in the training), iter = 20 (number of iterations over the training corpus). After training, we measured the similarity between documents in the training set by calculating the cosine similarity between their estimated vectors. We find that greater than 96% of the inferred documents register as most similar to themselves, which suggests the trained Doc2vec model is working in a usefully consistent manner. To provide face validity for our model, we randomly select a document and provide documents that register most and least similar (Supplementary Table 5). Using the trained model, we infer 100-dimensional vectors for each of the 45,553 articles contributed by 10,000 scholars randomly selected from name-disambiguated data. These vectors are used as an alternative measure for topics in the linear regression model introduced in the next section.

Predicting disruption percentile using multivariate linear regressions. From the WOS name-disambiguated data, we select all scholars with at least one cited paper published between 1954 and 2014. In this way, we obtain 38,000,470 scholars and 96,386,516 articles (articles are counted repeatedly if they appear across the publication records of different scholars). For each article, we construct five groups of variables for each article: (1) disruption percentile, (2) team size, or number of authors (this group includes 15 dummy variables for teams of size 2 to 15, with articles having 15 or more authors aggregated into a single 15+ variable and solo-authored papers as the reference category); (3) publication year; (4) topic ID, which is a categorical variable with ten values, covering gross topic areas ranging from the physical sciences to the social sciences (see Fig. 3b); and (5) author ID, a numeric index for the scholar with whom each paper is affiliated. We run two regression models, with and without author fixed effects. In both, we cluster standard errors using author ID. We use the `reghdfe` package³⁶ in STATA13 to run the regressions, which automatically identifies singleton groups and removes them

iteratively; the number of observations in the author fixed-effect model is therefore lower than that in the random effects model³⁶.

We also test an alternative measure of topics using a smaller sample of 10,000 randomly selected scholars. We construct 100 continuous variables varying from -3.2 to 3.2 that characterize topics for which we trained a Doc2vec model using titles and abstracts of 100,000 articles, as described above. We use this model to fit each of the 45,553 articles in the regression data and find that they provide results very similar to those using topic ID in predicting disruption percentile, which verifies the observed association between team size and disruption. Note that our linear regression models have some limitations. The name disambiguation algorithm used in this paper favours authors in large teams as similar sets of co-authors are used to help to disambiguate authors. Our approach also favours articles published in recent years, and those with active scholars that have more data and are therefore more easily identified.

Validating disruption. The disruption measure that we studied is calculated from citation networks. We conduct five independent analyses to validate and more richly characterize this structural definition of disruption. The first two investigations involve the alignment of disruption with expert notions of disruption and development; the second two tests link disruption to the process of self-consciously disrupting or developing the landscape of previous work; and the last inquiry characterizes the full range of expressed behaviours associated with disrupting or developing in science. Specifically, these include: (1) linking articles with Nobel prizes and showing that, across fields, Nobel-prize-winning papers registered among the most disruptive, which validates the notion that that expert assessment of path-breaking research systematically breaks the path of acknowledgement to former work upon which it builds; (2) fielding and analysis of an independent survey of scholars from a range of fields; this survey invited these scholars to propose disruptive and developing articles that confirmed our measure; (3) association of article type and disruptiveness, revealing that review articles—which explicitly summarize previous original research—are substantially more developmental than the work they review; (4) identification and analysis of informal eponymous references in article titles and abstracts that signalled how researchers in developing articles explicitly expressed their intention to build on important, prior research; and (5) extraction and analysis of distinguishing words that descriptively differentiate disruptive from developing articles. We detail each of these in the sections below.

Nobel Prize disruption. We evaluated the association between Nobel Prize award-winning articles (in a variety of fields) and disruption. The Nobel Prizes were established as a consequence of Alfred Nobel's last will and testament³⁷ drafted in 1895, which stated that the interest from his remaining fortune should be used to confer prizes on 'those who, during the preceding year, shall have conferred the greatest benefit to mankind ... to the person who shall have made the most important discovery or invention within the field of physics ... the most important chemical discovery or improvement ... the most important discovery within the domain of physiology or medicine ...'. In our sample of 877 papers directly connected with a Nobel prize (covering the time period from 1902 to 2009), the average disruption is 0.10, which ranks within the top 2% of all WOS papers from the same time period, selecting as control group 3,372,570 papers from the same 178 journals and years. This pattern is strong and substantial for prizes in Physiology or Medicine (316 papers), Physics (284 papers), and Chemistry (277 papers). Incidentally, we find that the probability of observing small-team, disruptive papers is nearly three times as high among Nobel-Prize-winning papers than those in the control group (Extended Data Fig. 10). This suggests that major scientific communities recognize work as important and path-breaking that has also been cited independently from the work upon which it builds, signifying a break in the path of acknowledgement.

Disruption survey. We fielded an open-ended survey, performed in person, over the telephone or using Skype, which was approved by the University of Chicago Institutional Review Board (IRB18-1248). The survey requested that scholars across different fields propose papers that either disrupt or develop science in their fields, anchoring those definitions with the following discussion: 'Developing papers represent extensions or improvements of previous theory, method or findings (note that many papers will extend some scientific elements, while keeping others the same). Disrupting papers represent punctuated advances beyond previous theory, methods or findings (note that almost no papers can successfully disrupt all scientific elements; if they disrupt some things, they likely develop and keep others the same)'. We then provided respondents with the BTW-model and Bose-Einstein condensation papers to demonstrate the kinds of papers that we would define as disruptive and developmental, and to demonstrate how developing papers could also be important. Respondents then proposed from three to ten disrupting and developing papers. Our panel of scientists were solicited from ten prominent research-intensive institutions across United States, China, Japan, France and Germany. These scientists had training that ranged across mathematics,

physics, chemistry, biology, medicine, engineering, computer science, psychology and economics. Among the 20 scholars from whom we received 190 answers, 100% of their proposals agreed with our measure for the most disruptive paper they mentioned according to our measure (and all but six of their proposals agreed with our measure for the most developing paper they mentioned). The average disruption of papers nominated as disruptive is 0.2147, among the top 2% of most disruptive papers. The average disruption of papers nominated to be developing is -0.011 , among the bottom 13% (Fig. 1c). This analysis resulted in an overall prediction area under the curve of 0.83, which suggests a predictive accuracy of 83% and a much stronger sensitivity to extremes. We present a selected list of disruptive papers in Supplementary Table 2.

Review articles versus original research. Review articles channel attention to important past work, and thus should systematically tend to be more developmental than disruptive. To test this hypothesis, we separate review articles from the original research articles they review by culling journals with the words both 'annual' and 'review' in the title, which resulted in a sample of 22,672 review articles published in 48 journals between 1954 and 2014. We compare these with the 1,338,808 articles reviewed (cited) by them. This reveals that reviews, which explicitly summarize previous original research, are substantially more developmental than original research. Precisely, the mean disruption for reviews (-0.0009) corresponds to the 46th percentile of the disruption distribution (based on all cited papers published between 1900 and 2014), and the non-review mean (0.0008) corresponds to the 77th percentile of the disruption distribution. This difference indicates that original research articles are much more likely to be disruptive than work that reviews them.

Informal reference. To further validate the link between our ex post measure of disruption with the search strategy in the original work that eventually comes to be received by the community as disruptive, we identified research that specifically signalled an intention to extend the important work of earlier authors by extracting all eponyms or informal references to prior authors' work in titles and abstracts, including 'Bose' and 'Einstein' from Bose-Einstein condensation, 'Bohr' from 'Bohr radius', 'Higgs' from 'Higgs boson' and so on. Specifically, we analyse 27,728,266 WOS articles between 1954 and 2014 with one or more citations, in which for each paper we construct a list of the family names for scholars who authored any of the papers cited in the references. We then identify whether these names also appeared in title or abstract. We found that nearly a million research papers in the Web of Science—0.61% (148,303) of titles and 3.0% (727,254) of abstracts—contain the names of previous authors or concepts and phenomena named after them. Articles that develop previous science, according to our future citation-based measure of disruption, are 250% more likely to reference former research by author name in title or abstract, which suggests an explicit intention to extend previous work and attract the attention of audiences that have appreciated it (Supplementary Table 1). This validates our measure by revealing its alignment with other rich signals of linkage to past science. It also confirms another core dimension of the disruption measure that was not discussed by the original authors of this measure: a creative work's future disruptive impact is strongly predicted by its search for the ideas upon which it will build. Research building on previous work that is either (1) sufficiently famous, such that it has been canonized, with the original author's name attached to the phenomena by the community, or (2) sufficiently recent, such that the author's name is familiar to the community, is much more likely to be received by that same community as an important extension of the prior known work. Our paper further suggests that this kind of developing activity is much more commonly performed by large teams.

Distinguishing words. Finally, we examined the titles and abstracts that introduce papers, which are eventually determined by our measure to be disruptive or developing papers. We identified those words that are most and least predictive of disruption, as measured by the relative ratio of their presence in disruptive versus developing articles. Specifically, we selected titles and abstracts from 24,174,022 WOS papers published between 1954 and 2014 with one or more citations for which $D \neq 0$. We assign them into two groups: disrupting articles ($D > 0$; 6,397,815) and developing articles ($D < 0$; 16,266,398). For words observed in both groups (1,033,879 words for titles and 3,492,223 words for abstracts), we calculate the ratio of their frequency in disrupting versus developing articles. In Supplementary Table 3, we present a sample of the most popular words with ratios that deviate significantly from 1. These distinguishing words (grouped by part of speech) characterize the manner in which articles come to disrupt or develop science. For example, 'technique', 'device', 'tool', and 'measure' are among the nouns that most distinguish titles and abstracts from disruptive articles. This suggests that new approaches are often used to disrupt science and technology with new findings and scientific possibilities. By contrast, 'theory', 'model' and 'hypothesis' are all significantly and strongly associated with articles that develop ideas from previous work. Verbs associated with disruptive article titles and abstracts include 'advance', 'introduce' and 'change', which suggests the introduction of new approaches and new causal forces to a scientific domain.

On the other hand, ‘endorse’, ‘confirm’ and ‘demonstrate’ are much more likely to be found in developing articles. These focus on confirming and incrementally altering existing scientific components. For example, the contrast between ‘introduce’ and ‘confirm’ is consistent with our definitions of disrupting and developing papers in terms of whether questions are asked or solved, respectively. Finally, adverbs and prepositions that distinguish disruptive article titles and abstracts include questioning words (who, why, what, where and when), which provides additional support for the increased likelihood of disruptive research to pose new questions. By contrast, ‘during’, ‘after’ and ‘from’ characterize work that develops—and integrates—insights from previous investigations. These distinguishing words are highly suggestive regarding strategies that characterize disruptive work. Moreover, they highlight differences in the search and positioning of ideas that come to correlate with how those articles are received, which forms the basis of our disruption measure.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

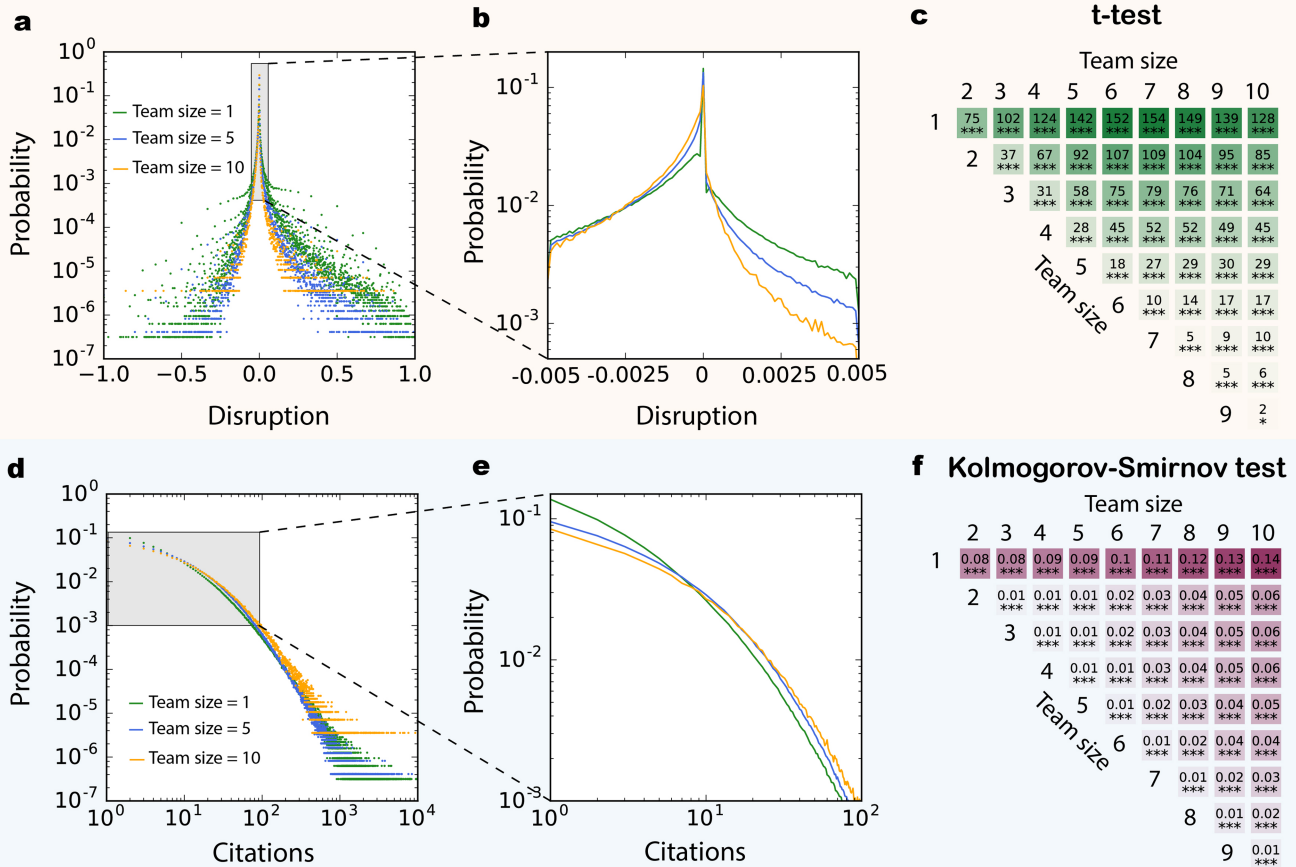
Code availability. All code is available at <http://lingfeiwu.github.io/smallTeams>.

Data availability

Data are available at <http://lingfeiwu.github.io/smallTeams>. Other related, relevant data are available from the corresponding author upon reasonable request.

32. Alcácer, J., Gittleman, M. & Sampat, B. Applicant and examiner citations in U.S. patents: an overview and analysis. *Res. Policy* **38**, 415–427 (2009).
33. Schulz, C., Mazloumian, A., Petersen, A. M., Penner, O. & Helbing, D. Exploiting citation networks for large-scale author name disambiguation. *EPJ Data Sci.* **3**, 11 (2014).
34. Mutz, R., Bornmann, L. & Daniel, H.-D. Cross-disciplinary research: What configurations of fields of science are found in grant proposals today? *Res. Eval.* **24**, 30–36 (2015).
35. Le, Q. & Mikolov, T. Distributed representations of sentences and documents. In *Proc. 31st International Conference on Machine Learning* (eds Xing, E. P. & Jebara, T.) 1188–1196 (PLMR, Beijing, 2014).
36. Correia, S. A feasible estimator for linear models with multi-way fixed effects. Preprint at <http://scorreia.com/research/hdfe.pdf> (2016).
37. Full text of Alfred Nobel’s will, available at <https://www.nobelprize.org/alfred-nobel/full-text-of-alfred-nobels-will/> (accessed 25 September 2018).

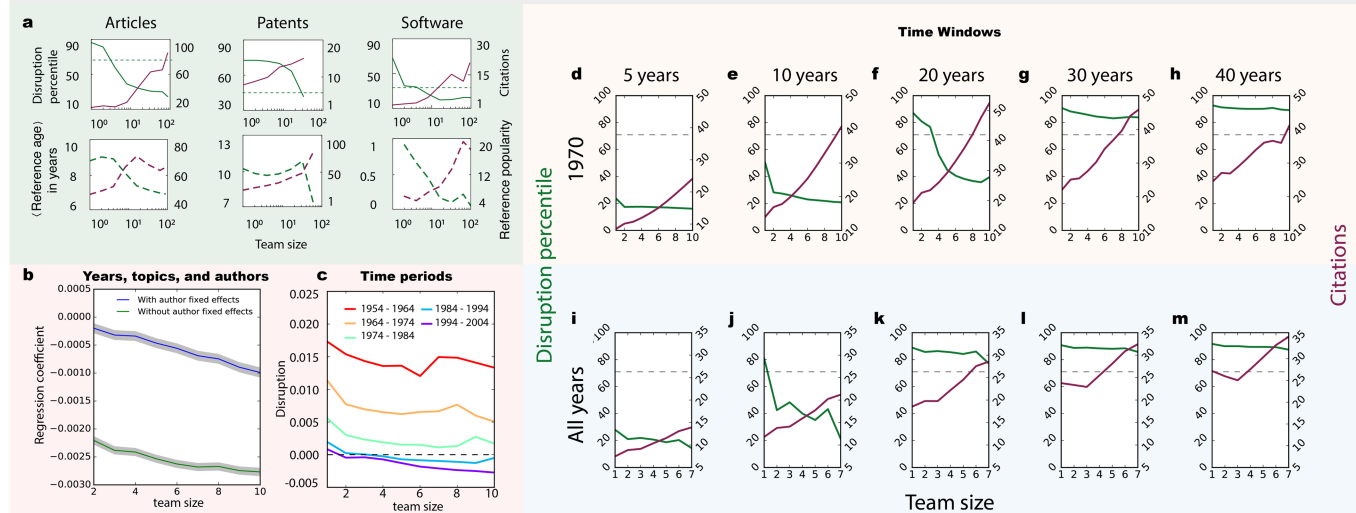
Comparing Citation and Disruption Distributions across Team Sizes



Extended Data Fig. 2 | Comparing citation and disruption distributions across team sizes. We select 27,728,266 WOS papers of at least one citation published between 1954 and 2014. **a, b**, The distribution of disruption changes with team size (**a**); magnified versions of the grey area shown in **b**. **c**, We test differences in the distribution of disruption between each pair of team sizes from one to ten using a two-sample *t*-test. The *t*-statistics are given in green cells and the darkness of green is proportional to the size of each *t*-statistic. Asterisks under the numbers indicate *P* values. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$. All pairs of tested disruption distributions significantly differ from one another. **d, e**, The

distribution of citation changes with team size (**d**); magnified versions of the grey area shown in **e**. All figures clearly demonstrate how small teams oversample more disruptive and less impactful work. **f**, We test differences in the distribution of citations between team sizes using two-sample Kolmogorov-Smirnov tests, which are recommended for long-tailed distributed data. Numbers in cells show Kolmogorov-Smirnov statistics and the underlying asterisks indicate *P* values. All pairs of the tested citation distributions significantly differ from one another. Comparing disruption distributions with the Kolmogorov-Smirnov test reveals the same patterns of difference.

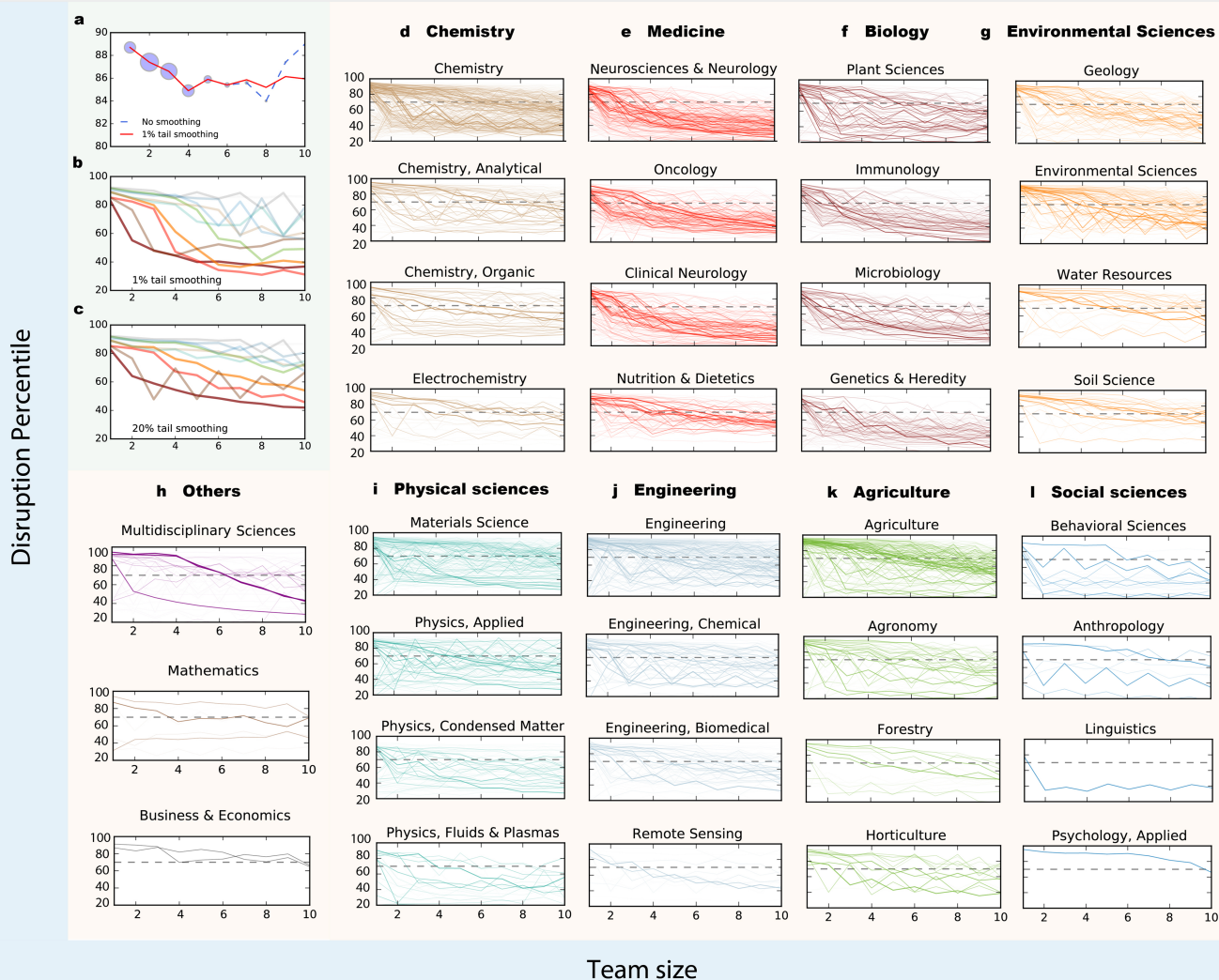
Small Teams Disrupt: Scaling Up Team Size and Controlling for Years, Topics, Authors, Time Periods, and Time Windows



Extended Data Fig. 3 | Decreasing disruption is robust across years, topics, authors, time periods and windows of disruption. **a**, For research articles (24,174,022 WOS articles published between 1954 and 2014), patents (2,548,038 US patents assigned between 2002 and 2014) and software (26,900 GitHub repositories uploaded between 2011 and 2014), median citations (red curves, indexed by right y axis) increase with team size from 1 to 100 (rather than 1 to 10 as in Figs. 2a–c, 4a–c), whereas the average disruption percentile (green curves, indexed by left y axis) decreases with team size. For all datasets, we present work with one or more citations. Green dotted lines show the point at which $D = 0$, the transition from development to disruption. Bootstrapped 95% confidence intervals are shown as grey zones. **b**, Plot of the regression coefficients of disruption (rather than disruption percentile as in Fig. 3c) on team size, from linear regressions controlling for publication year, topics and author. The regression is based on the 96,386,516 WOS research articles (articles

are counted repeatedly if they appear across the publication records of different scholars) contributed by 38,000,470 name-disambiguated scholars. **c**, The negative correlation between disruption and team size holds across time periods. In contrast to the main body of the paper, which renders disruption in terms of percentile change, here we measure it in the native metric of disruption to highlight the shift with time. Earlier cohorts (red curves) are more disruptive than later cohorts. Nevertheless, with changes in team size, each cohort of papers traverses a majority of the total variation of disruption for that cohort. **d–h**, Decreasing disruption percentile and increasing citations with growing team size are robust to changes in the width of the time-window of observation from 5 years to 40 years for 166,310 WOS articles published in 1970. **i–m**, As in **d–h**, but using 24,174,022 WOS papers published between 1954–2014; we observe the same pattern.

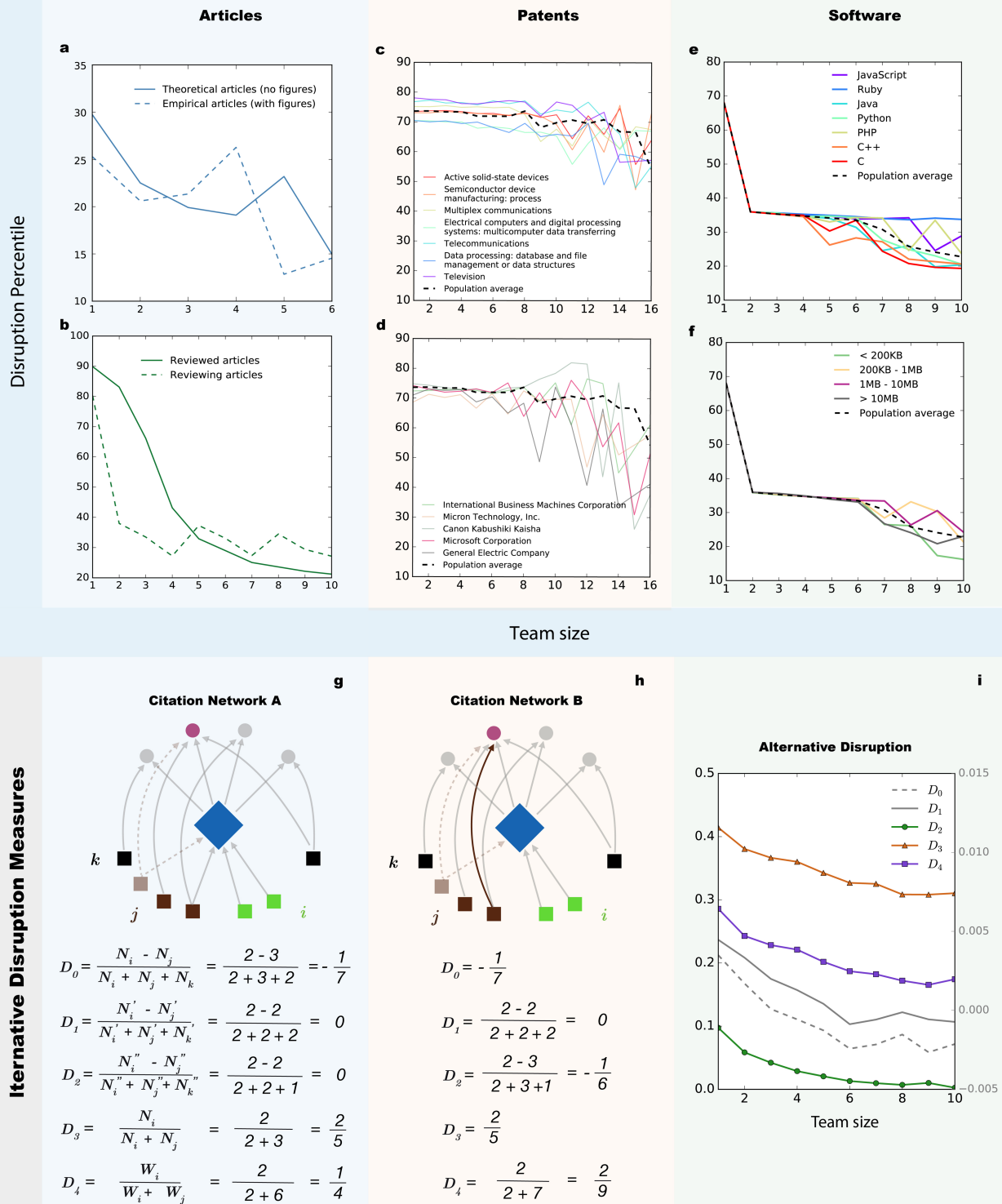
Small Teams Disrupt: Controlling for Journals



Extended Data Fig. 4 | Decreasing disruption is robust when controlling for journal. **a–c**, Weighted moving average technique for data smoothing. The relationship between team size and disruption may be noisy owing to lack of data when we analyse WOS articles from the same journal. As shown in **a**, less than 1% of articles in ‘Artificial intelligence’ (a subfield of ‘Computer and Information Technology’) have more than six authors, but these articles contribute to substantial variance in the data. We use the moving average technique to limit noise in the data. More specifically, we define a parameter k , which provides the threshold value of m_k for team size m such that $P(m > m_k) < k$. For any data point with a team size greater than m_k , its disruption percentile DP_m is updated to be the average between its current value and the value of its left neighbour, DP_{m-1} , weighted by corresponding sample sizes (the number of articles for a given team size). Panel **a** shows curves for

the subfield ‘Artificial Intelligence’ before (blue dashed curve) and after (red curve) smoothing, in which the size of blue circles is proportional to sample size. Panels **b** and **c** show how smoothing depends on the value of k across ten randomly selected subfields. In **d–l**, each curve corresponds to a journal (only journals with more than three data points are shown) and each panel corresponds to a subfield. There are 15,146 journals, 258 subfields and 10 major fields represented in our WOS data. Owing to the limited figure size, only four subfields are shown for each field. Curves are smoothed by setting the smoothing parameter $k = 0.2$. The darkness of curves is equally proportional to sample size and the absolute value of the regression coefficient examining the impact of disruption percentile on team size, such that journals with more articles and that display stronger (both negative and positive) relationships are more distinguishable from the background.

Small Teams Disrupt: Controlling for Task, Institution, Platform, and Project Scale



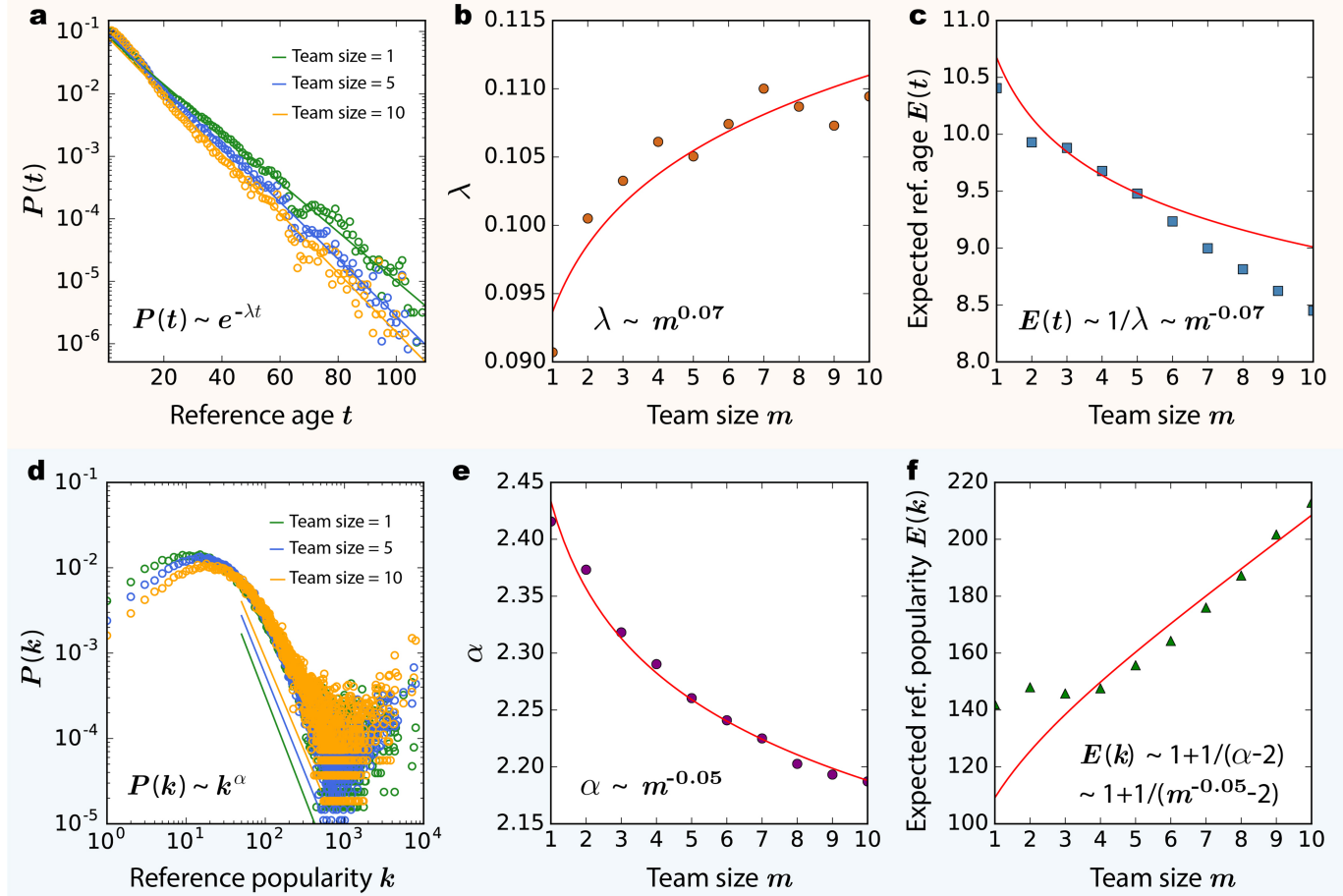
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Decreasing disruption is robust when controlling for task, institution, platform, project scale and alternative disruption measures.

a, b, Comparison between theoretical and empirical articles (**a**) and review and non-review articles (**b**). **a,** We separate 4,258 papers from www.arXiv.org published between 1992 and 2003 into two groups on the basis of the number of figures they contain; this grouping comprised 1,502 articles without figures and 2,756 articles with figures. The assumption is that empirical papers tend to contain more figures than theoretical papers²³. We match these articles to the WOS datasets and observe that for both theoretical and empirical articles, the disruption percentile decreases with the growth of team size. **b,** We select two groups of WOS articles on the basis of journal name; 22,672 reviewing articles published across 48 journals that have both 'annual' and 'review' in the title, and their 1,338,808 references (reviewed articles). For both reviewing and reviewed articles, the disruption percentile decreases with team size. **c, d,** Comparison of US patents across classes and owners. We plot the disruption percentile against team size for the seven most popular classes of patents (92,175 patents) (**c**) and the top five companies legally assigned the most patents (21,261 patents) (**d**) from 2002 to 2009. We observe that the decrease in disruption and increase in team size holds broadly across classes and owners. The moving average technique used in Extended Data Fig. 4 is used to smooth the curve (smoothing parameter $k = 0.1$). As sample size decreases rapidly with team size in the patent data, we assigned equal weights across team sizes in applying the smoothing technique. **e, f,** Comparison of GitHub software projects across programming languages and code-base sizes. We plot the disruption percentile against team size for the seven most popular programming languages (18,702

(**e**) and four scales of code-base sizes (24,853 code-bases) (**f**) from 2011 to 2014. The decrease in disruption with growth of team size holds broadly across programming languages and code-base sizes. **g,** Simplified citation networks comprising focal papers (blue diamonds), references (grey circles) and subsequent work (rectangles). Subsequent work may cite: (1) only the focal work (*i*, green), (2) only its references (*k*, black) or (3) both focal work and references (*j*, brown). A reference identified as popular is coloured in red, and self-citations are shown by dashed lines (with corresponding subsequent work coloured in light brown). Five definitions of disruption are provided for comparison. D_0 is the definition of disruption used in the main text. D_1 is defined the same way as D_0 , but with self-citations excluded. D_2 is defined the same way as D_0 , but only considers popular references. We identified references as popular that received citations within the top quartile of the total citation distribution (≥ 24 citations). D_3 simplifies D_0 by only measuring the fraction of papers that cite the focal paper and not its references, among all papers citing the focal paper, which equals $n_i/(n_i + n_j)$. D_4 is similar to D_3 , but considers the number of citations and not papers cited in calculating the fraction (for example, if a single referenced paper is cited five times, then it receives a count of five rather than one in this measure). **h,** A citation network copied from **g**, with one additional citation edge (brown curve) added. As a consequence, some—but not all—disruption measure variants change. **i,** All disruption measures decrease with team size. D_0 and D_1 are indexed by the right y axis and other disruption measures are indexed by the left y axis. One hundred thousand randomly selected WOS papers (97,188 papers remained after excluding missing data) are used to calculate these disruption values.

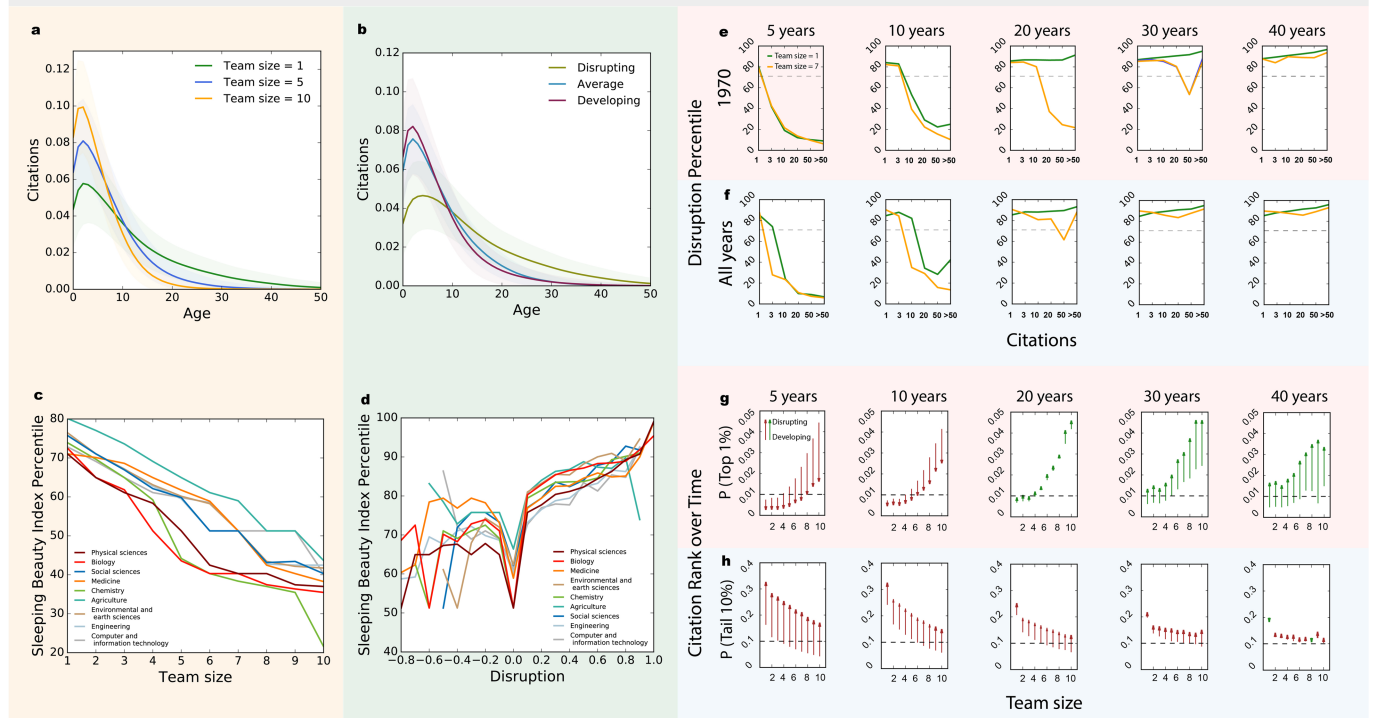
Small Teams Cite Earlier and Less Popular References



Extended Data Fig. 6 | Small teams cite earlier and less-popular references. **a**, We select 1,127,518 WOS articles published in 2010 and find that the probability of observing reference j of age t decreases exponentially with t , such that $P(t) \sim e^{-\lambda t}$. For larger teams $P(t)$ decreases faster with t , suggesting that λ is determined by team size m . **b**, The relationship between m and λ (orange circles) can be fitted as $\lambda \sim m^{0.07}$ (red curve). **c**, From **a** and **b**, we can derive the dependency of $E(t)$, the expected value of t , on m by integrating $P(t)$ from zero to maximum t . This gives $E(t) \sim 1/\lambda \sim m^{-0.07}$. Empirical data (blue rectangles) are consistent with this prediction (red curve). **d**, Probability of observing reference

j with k citations decreases with k , supporting the relationship $P(k) \sim k^{-\alpha}$. To control the time window, we include only references published in 2005. For larger teams $P(k)$ decreases more slowly with k , suggesting that α is affected by m . **e**, The empirical relationship between m and α (purple circles) and the fitting function as $\alpha \sim m^{-0.05}$ (red curve). **f**, From **d** and **e**, we can derive the dependency of $E(k)$, the expected value of k , on m by integrating $P(k)$ from minimum to maximum k . This gives $E(k) \sim 1 + 1/(\alpha - 2) \sim 1 + 1/(m^{-0.05} - 2)$. The empirical data (green triangles) are consistent with this prediction (red curve).

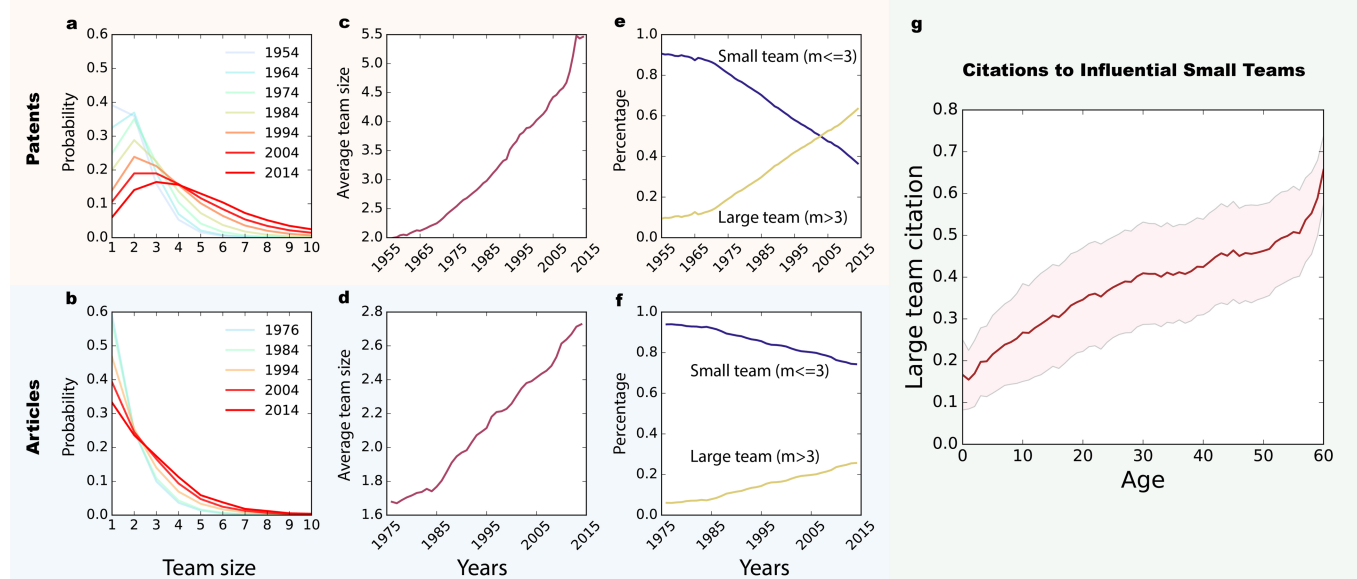
Citation Delay to Small and Disrupting Teams

**Extended Data Fig. 7 | Citation delay to small and disrupting teams.**

a, b, The decay of citations to WOS articles changes with team size and disruption. We selected 95,474 papers with 200–300 citations from 1954 to 2014, and plot the probability of being cited against article age. Longer delays in citation are observed in smaller (**a**) and more disrupting (**b**) teams. In **b**, purple (37,805 papers), blue (4,931 papers) and green (26,698 papers) curves correspond to 0–10, 55–65 and 90–100 percentiles of disruption, respectively. In both panels, curves are smoothed by a running average with a time window of five years. The coloured area shows one standard deviation of these averages. **c, d,** The Sleeping Beauty index²⁴ captures a delayed burst of attention by calculating convexity in the citation distribution of a particular work over time. The index is highest when a paper is not cited for some substantial period before receiving its maximum (which corresponds to belated appreciation), zero if the paper is cited linearly in the years following publication, and negative if citations chart a concave function with time (which traces early fame diminishing thereafter). We observe that the Sleeping Beauty index percentile decreases

markedly with team size (**c**) and increases with disruption (**d**) across fields. **e, f,** The negative correlation between disruption percentile and impact in the short term (within 10 years) turns positive in the long term (over 30 years) for the 166,310 papers published in 1970 (**e**). The same pattern is observed when all 22,174,022 papers from 1954 to 2014 are used (**f**). **g, h,** Achieving substantial citation attention for disruptive work occurs over the long term, if at all, whereas the risk of failure from disruption occurs over both the short and long term. Arrows trace the distance between the mean of future citation success (**g**) or failure (**f**) from developing to disrupting work produced by teams of each specified size. The probability of becoming one of the top 1% most-cited articles is higher for developing teamwork (negative disruption, the origin of arrows) within 20 years and higher for disrupting teamwork (positive disruption, the target of arrows) over 30 years across team sizes (**g**). The probability of becoming one of the tail 10% least-cited articles is almost always higher for disrupting teamwork than developing teamwork across team sizes and time windows (**h**).

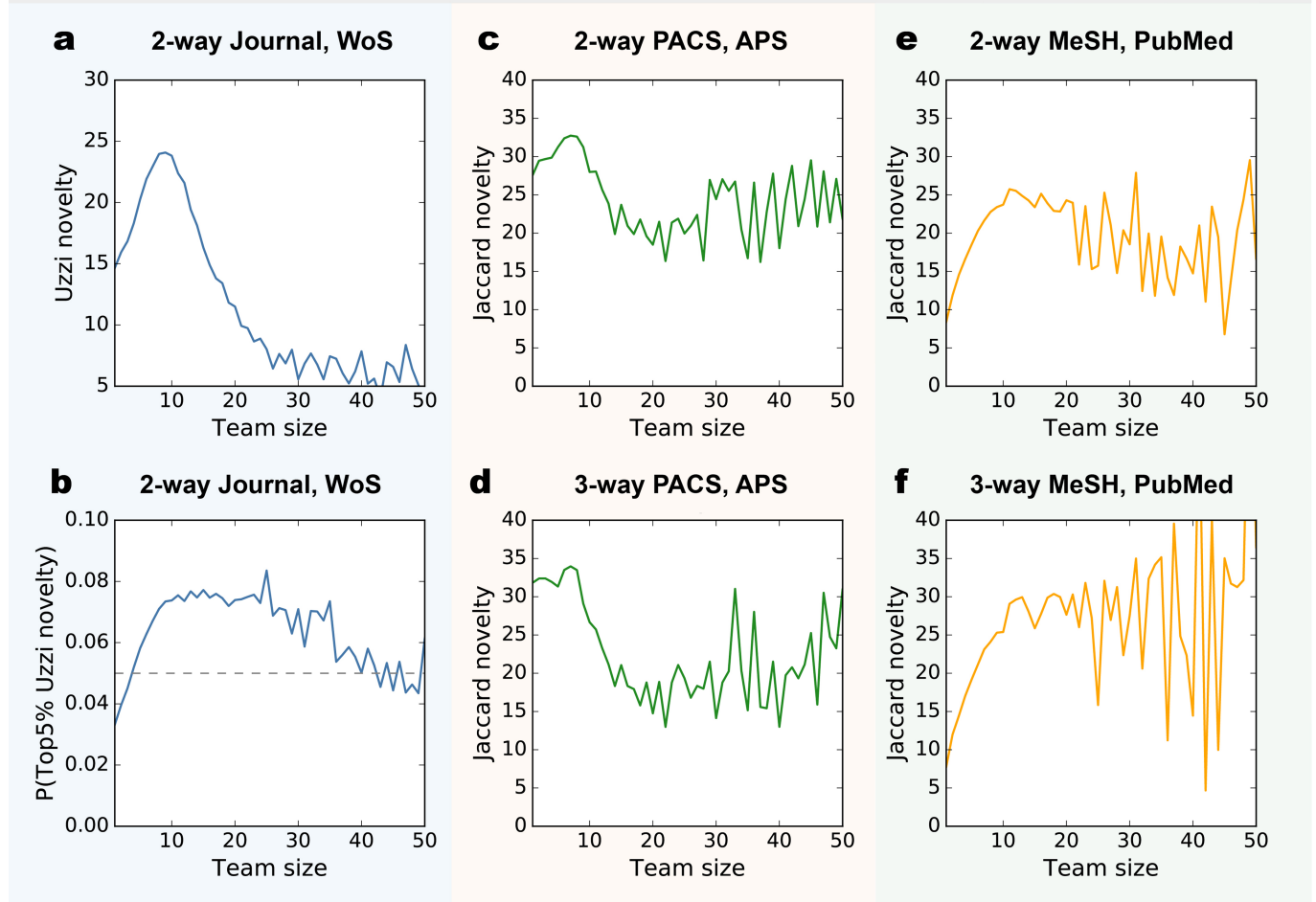
The Ripple Effect of Shrinking Small Team Population



Extended Data Fig. 8 | The ripple effect of a shrinking small team population. **a–f**, The decline of small teams. **a**, **b**, Evolution of team-size distributions over time for WOS articles (**a**) and US patents (**b**). The distributions skew towards large teams over time. **c**, **d**, Average team size of articles increased from 2.0 to 5.5 between 1954 and 2014, and for patents team size increased from 1.7 to 2.7 between 1976 and 2014. **e**, **f**, Percentage of small teams (in which the number of team members $m \leq 3$) decreased from 91% to 37% for articles, and from 94% to 74% for patents during the

period of observation. **g**, The ripple effect. We select 2,640 small teams ($m \leq 3$) from WOS articles that are among the top 1% in number of citations they received, as well those among the top 1% within the Sleeping Beauty index distribution²⁴. We analyse the citations to these articles and find that the fraction of large teams ($m > 3$) increases over time. The red curve shows the average fraction of citations from large teams and the pink area spans one standard deviation. The selected 2,640 small-team articles are eventually cited by 657,946 large-team articles.

Diseconomies of Scale in Combinatorial Novelty

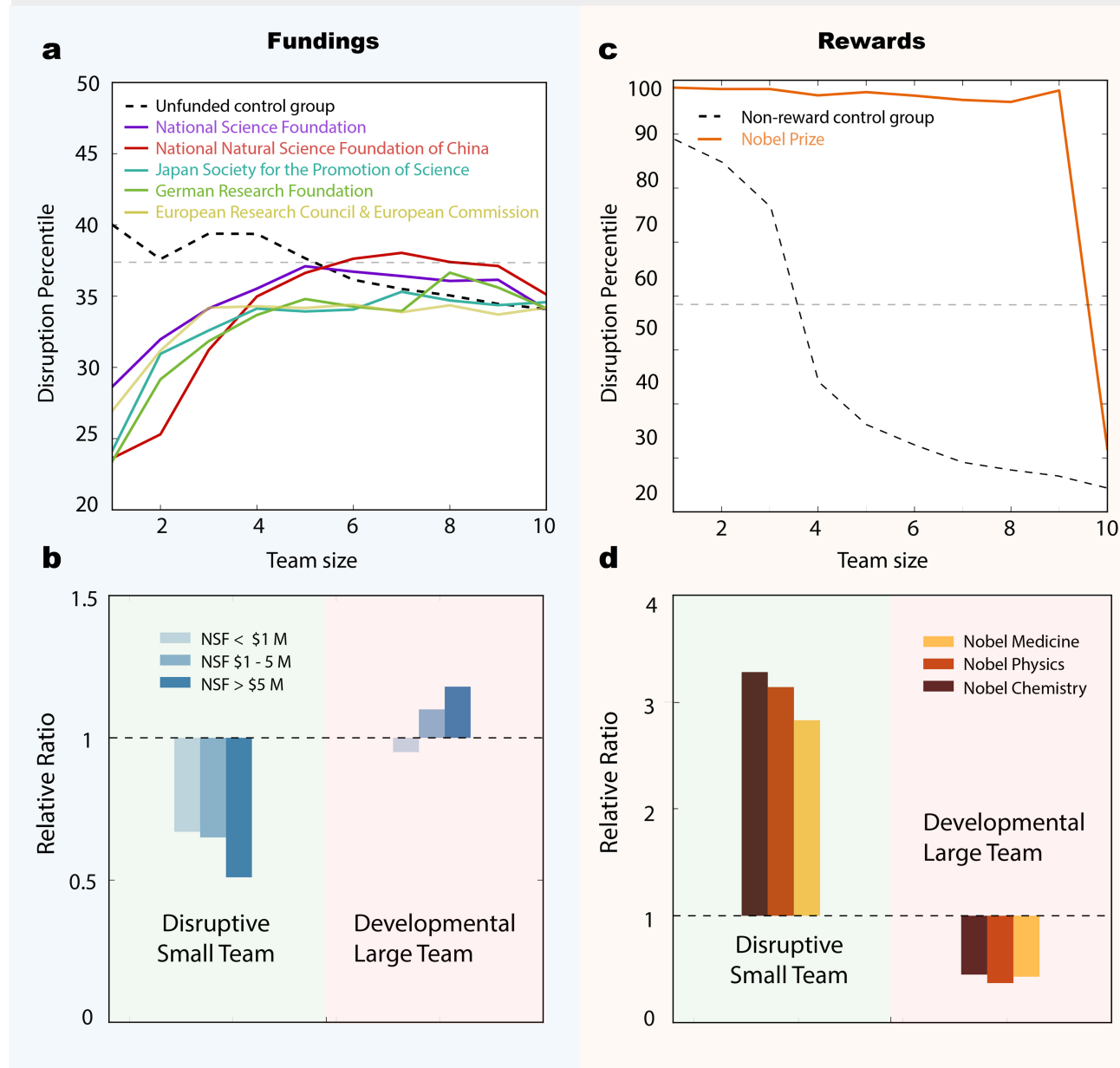


Extended Data Fig. 9 | Diseconomies of scale in combinatorial novelty.

a, b, Changes in journal-based combinatorial novelty with team size from WoS articles. We calculate the pairwise combinatorial novelty of journals in the references of an article using a previously published novelty measure²⁸. This novelty measure is computed as the tenth percentile value of z -scores for the likelihood that reference sources combine, so a lower value of this index indicates higher novelty²⁸. Here we convert this measure to percentiles and subtract from 100 to improve readability, such that a higher score indicates greater novelty. It seems natural that a larger team would provide access to a wider span of literature. We find that novelty does increase with team size, but with diminishing marginal increases to novelty with each additional team member. Beyond a team size of ten, novelty decreases sharply (**a**). The probability of observing papers within the top 5% of the novelty distribution increases, and then decreases, with team size. The dotted line shows the null model that the

probability of high novelty is invariant to team size (**b**). **c, d,** Calculation of combinatorial novelty in a different way. We select 241,648 papers published in American Physical Society Journals, 1990–2010, and analyse the probability of two-way (pairwise) and three-way combinations of the ‘Physics and Astronomy Classification Scheme’ codes using the Jaccard index. Similar to the novelty measure used in **a** and **b**, in the Jaccard index a lower value indicates higher novelty; we therefore convert it into percentiles and subtract from 100 such that a higher score indicates greater novelty. Again, we observe diminishing marginal increases to novelty with the growth of team size. **e, f,** We select 8,232,630 PubMed papers from between 1990 and 2010 and analyse the probability of two-way and three-way combinations of medical subject headings using Jaccard indices. The diminishing marginal increases to novelty effect are also observed in this context.

Small, Disrupting Teams are Awarded but Underfunded



Extended Data Fig. 10 | Small, disruptive teams contribute disproportionately to Nobel Prizes and are underrepresented with government funding. **a**, Underfunded small-team, disruptive research. Disruption percentile versus team size for WOS papers either not annotated as funded, or as funded by the largest government agencies around the world. The 477,702 funded papers cover the time period 2004–2014, and include 198,103 for NSF, 80,448 for NSFC, 81,296 for ERC and EC, 75,881 for DFG and 58,275 for JSPS. These papers are published across 7,325 journals, and a paper may be funded by multiple agencies. The average disruption of these papers is -0.0024 , ranking in the tail 31.0% of all WOS papers in the same period. We select 5,305,534 papers without any funding annotations from the same 7,325 journals and same time period (2004–2014) as a control group (dashed curve). The dashed grey line shows the mean disruption percentile for the control group. **b**, We select 191,717 papers published between 2008 and 2014 that acknowledged NSF with a grant number and retrieved grant size from the NSF website,

including 140,972 papers for less than or equal to 1 million US dollars, 24,370 papers for 1–5 million US dollars and 26,375 papers for more than 5 million US dollars. The green and red zones mark two regions of interest: small-team (three or fewer members) disruptive (positive disruption) papers in green and large-team developing work in red. The probability of observing small-team disruptive papers in NSF granted papers is almost half that of observing them in the control group. **c**, We select 877 Nobel-Prize-winning papers that cover the time period 1902–2009, including 316 papers in Physiology or Medicine, 284 papers in Physics and 277 papers in Chemistry. We select 3,372,570 papers from the same 178 journals and same time period (1902–2009) as a control group (dashed curve). The average disruption of the Nobel-prize-winning papers is 0.10, ranking among the top 2% of all WOS papers from the same period. **d**, The probability of observing small-team disruptive papers is nearly three times as high in Nobel-Prize-winning papers as in the control group.

Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) 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

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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

We performed most analysis with Python 3, using pandas dataframes and models, as described in the manuscript and Supplement. For the regression models, we used and Stata/SE 13.0

Data analysis

We performed all analysis with standard algorithms and data. We will also (redundantly) make our particular implementation of well-known code available in a public GitHub repository that indexed in the manuscript and supplement to maximize reproducibility.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

Our data involves use of public data including the GitHub public repository collection and the US Patent and Trademark office patent database. While we cannot redistribute these, we can and will publicly share how to access these resources and code regarding how to do it most effectively. We will also share all of the Web of Science data required to reproduce our analyses and figures.

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](https://www.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	Data on teams size were collected for tens of millions of productive teams, attributes of their productive output and its influence on future science and technology. This data was largely quantitative in nature, including author/inventor/developer number for each team, and network data regarding both how these objects searched through the space of past science and technology, and how their work was received by future generations.
Research sample	Our sample involved more than 65 million teams producing science publications, technology patents, and software. This ranged from the end of the 19th Century for articles, from 1976 until the present for patents and over the 21st Century for software products. All of these details are specified clearly in the manuscript and supplement. We also included all data available in scientific papers, technology patents, and software relevant prior to the works in question to evaluate search, and posterior to them to evaluate impact, disruption and its delay. These works represent a population of relevant artifacts and should not be viewed as a sample of some different populations. We do find, however, that all subsamples of the data confirm the pattern we see in the populations as a whole. Moreover, because the patterns we evaluate are consistent across these massive populations, we suggest that they have likely relevance to other contexts of science, technology and cultural production as well
Sampling strategy	We used all available data for our analysis of teams' search strategies, impact and disruption. We also subset the data and analyze it separately for subsamples, presenting results in the Supplement.
Data collection	Our data was collected through administrative procedures that archive journal articles, publicly serve patents, and facilitate the sharing of code.
Timing	The timing of our data collection involved collecting data through 2015, but only analyzing team work from years before this time, to allow time for the accumulation of citations, critical for our measurements.
Data exclusions	No data were excluded from the analysis, other than by administrative convention. For example, the US patent system did not collect its data digitally until 1976. Despite image data from US patents being available through the end of the 18th Century, we used the digitally native patent data, from 1976 until the present.
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

Methods

n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Most data from human participants was passively collected and curated through publicly available data. We also performed a small survey of scholars to solicit their nominations for most disruptive and developmental articles.
----------------------------	--

Recruitment

For the disruption validation survey, we assembled a panel of young scientists from ten prominent research-intensive institutions across U.S., China, Japan, France, and Germany who responded to an online solicitation request. Relevant fields of these scientists covered math, physics, chemistry, biology, medicine, engineering, computer science, psychology, and economics. Among the 20 scholars, we received 190 answers

Ethics oversight

University of Chicago Institutional Review Board (#IRB181248)

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