

# Papers and patents are becoming less disruptive over time

Michael Park,<sup>1</sup> Erin Leahey<sup>2</sup> Russell J. Funk,<sup>1\*</sup>

---

<sup>1</sup>Carlson School of Management, University of Minnesota,  
321 19th Ave S, Minneapolis, MN 55455, USA

<sup>2</sup>Department of Sociology, University of Arizona  
1145 E South Campus Drive, Tucson, AZ 85721, USA

\*To whom correspondence should be addressed; e-mail: rfunk@umn.edu.

**Summary** Theories of scientific and technological change view discovery and invention as endogenous processes<sup>1,2</sup>, wherein prior accumulated knowledge enables future progress by allowing researchers to, in Newton’s words, “stand on the shoulders of giants”<sup>3–7</sup>. Recent decades have witnessed exponential growth in the volume of new scientific and technological knowledge, thereby creating conditions that should be ripe for major advances<sup>8,9</sup>. Yet contrary to this view, studies suggest that progress is slowing in several major fields<sup>10,11</sup>. Here, we analyze these claims at scale across 6 decades, using data on 45 million papers and 3.9 million patents from 6 large-scale datasets, together with a novel quantitative metric—the CD index<sup>12</sup>—that characterizes how papers and patents change networks of citations in science and technology. We find that papers and patents are increasingly less likely to break with the past in ways that push science and technology in new directions. This pattern holds universally across fields and is robust across multiple different citation- and text-based metrics. Subsequently, we link this decline in disruptiveness to a narrowing in the use of prior knowledge, allowing us to reconcile the patterns we observe with the “shoulders of giants” view. We find that the observed declines are unlikely to be driven by changes in the quality of published science, citation practices, or field-specific factors. Overall, our results suggest that slowing rates of disruption may reflect a fundamental shift in the nature of science and technology.

While the past century witnessed an unprecedented expansion of scientific and technological knowledge, there are concerns that innovative activity is slowing<sup>13–15</sup>. Studies document declining research productivity in semiconductors, pharmaceuticals, and other fields<sup>10,11</sup>. Papers, patents, and even grant applications have become less novel and less likely to connect disparate areas of knowledge, both of which are precursors of innovation<sup>16,17</sup>. The gap between the year of discovery and the awarding of a Nobel Prize has also increased<sup>18,19</sup>, suggesting that today’s contributions do not measure up to the past. These trends have attracted increasing attention from policymakers, as they pose significant threats to economic growth, human health and well being, and national security, along with global efforts to combat grand challenges like climate change<sup>20,21</sup>.

Numerous explanations for this slowdown have been proposed. Some point to a dearth of “low hanging fruit” as the readily available productivity-enhancing innovations have already been made<sup>14,22</sup>. Others emphasize the increasing burden of knowledge; scientists and inventors require ever more training to reach the frontiers of their fields, leaving less time to push those frontiers forward<sup>13,23</sup>. Yet much remains unknown, not merely about the causes of slowing innovative activity, but also the depth and breadth of the phenomenon. The decline is difficult to reconcile with centuries of observation by philosophers of science, who characterize the growth of knowledge as an endogenous process, wherein prior knowledge enables future discovery, a view captured famously in Newton’s observation that if he had seen further, it was by “standing on the shoulders of giants.”<sup>3</sup> Moreover, to date, the evidence pointing to a slowdown is based on studies of particular fields, using disparate and domain-specific metrics<sup>10,11</sup>, making it difficult

to know whether the changes are happening at similar rates across areas of science and technology. Little is also known about whether the patterns seen in aggregate indicators may mask differences in the degree to which individual works push the frontier.

We address these gaps in understanding by analyzing 25 million papers (1945-2010) in the Web of Science (see Methods, “WoS data”) and 3.9 million patents from (1976-2010) in the United States Patent and Trademark Office’s Patents View database (see Methods, “Patents View data”). The WoS data include 390 million citations, 25 million paper titles, and 13 million abstracts. The Patents View data include 35 million citations, 3.9 million patent titles, and 3.9 million abstracts. Subsequently, we replicate our core findings on 4 additional datasets—JSTOR, the American Physical Society corpus, Microsoft Academic Graph, and PubMed—encompassing 20 million papers. Using these data, we join a novel citation-based measure<sup>12</sup> with textual analyses of titles and abstracts to understand whether papers and patents forge new directions over time and across fields.

## Measurement of disruptiveness

To characterize the nature of innovation, we draw on foundational theories of scientific and technological change<sup>2,24,25</sup>, which distinguish between two types of breakthroughs. First, some contributions improve existing streams of knowledge, and therefore consolidate the status quo. Kohn & Sham (1965)<sup>26</sup>, a Nobel-winning paper used established theorems to develop a method for calculating the structure of electrons, which cemented the value of prior research. Second, some contributions disrupt existing knowledge, rendering it obsolete, and propelling science and technology in new directions. Watson & Crick (1953)<sup>27</sup>, also a Nobel winner, introduced a model of the structure of DNA that superseded previous approaches (e.g., Pauling’s triple helix). Kohn & Sham and Watson & Crick were both important, but their implications for scientific and technological change were different.

We quantify this distinction using a measure—the CD index<sup>12</sup>—that characterizes the consolidating or disruptive nature of science and technology (Figure 1). The intuition is that if a paper or patent is disruptive, the subsequent work that cites it is less likely to also cite its predecessors; for future researchers, the ideas that went into its production are less relevant (e.g., Pauling’s triple helix). If a paper or patent is consolidating, subsequent work that cites it is also more likely to cite its predecessors; for future researchers, the knowledge upon which the work builds is still (and perhaps more) relevant (e.g., the theorems Kohn & Sham used). The CD index ranges from -1 (consolidating) to 1 (disruptive). We measure the CD index five years after the year of each paper’s publication (indicated by  $CD_5$ , see Extended Data Figure 1 for the distribution of  $CD_5$  among papers and patents and Extended Data Figure 2 for analyses using alternative windows<sup>28</sup>). For example, Watson & Crick and Kohn & Sham both received over a hundred citations within five years of being published. However, the Kohn & Sham paper has a  $CD_5$  of -0.22 (indicating consolidation), whereas the Watson & Crick paper has a  $CD_5$  of 0.62 (indicating disruption). The CD index has been

validated extensively in prior research, including through correlation with expert assessments<sup>12,29</sup>.

## Declining disruptiveness

Across fields, we find that science and technology are becoming less disruptive. Figure 2 plots the average  $CD_5$  over time for papers (Figure 2a) and patents (Figure 2b). For papers, the decrease between 1945 and 2010 ranges from 91.9% (where the average  $CD_5$  dropped from 0.52 in 1945 to 0.04 in 2010 for the Social Sciences) to 100% (where the average  $CD_5$  decreased from 0.36 in 1945 to 0 in 2010 for the Physical Sciences); for patents, the decrease between 1980 and 2010 ranges from 78.7% (where the average  $CD_5$  decreased from 0.30 in 1980 to 0.06 in 2010 for Computers and Communications) to 91.5% (where the average  $CD_5$  decreased from 0.38 in 1980 to 0.03 in 2010 for Drugs and Medical). For both papers and patents, the rates of decline are greatest in the earlier parts of the time series, and for patents, they appear to begin stabilizing between the years 2000 and 2005. For papers, since about 1980, the rate of decline has been more modest in the Life Sciences and Biomedicine and the Physical Sciences, and most dramatic and persistent in the Social Sciences and Technology. Overall, however, relative to earlier eras, recent papers and patents do less to push science and technology in new directions. The general similarity in trends we observe across fields is noteworthy in light of “low hanging fruit” theories<sup>14,22</sup>, which would likely predict greater heterogeneity in the decline, as it seems unlikely fields would “consume” their low hanging fruit at similar rates/times.

## Linguistic change

The decline in disruptive science and technology is also observable using alternative indicators. Because they create departures from the status quo, disruptive papers and patents are likely to introduce new words (e.g., words used to create a new paradigm might differ from those that are used to develop an existing paradigm<sup>30,31</sup>). Therefore, if disruptiveness is declining, we would expect a decline in the diversity of words used in science and technology. To evaluate this, Figures 3a and 3d document the type-token ratio (i.e., unique/total words) of paper and patent titles over time (see Supplementary Information Section 1). We observe substantial declines, especially in the earlier periods, prior to 1970 for papers and 1990 for patents. For paper titles (Figure 3a), the decrease (1945-2010) ranges from 76.5% (Social Science) to 88% (Technology); for patent titles (Figure 3d), the decrease (1980-2010) ranges from 32.5% (Chemical) to 81% (Computer and Communications). For paper abstracts (Extended Data Figure 3a), the decrease (1992-2010) ranges from 23.1% (Life Science and Biomedicine) to 38.9% (Social Science); for patent abstracts (Extended Data Figure 3b), the decrease (1980-2010) ranges from 21.5% (Mechanical) to 73.2% (Computers and Communications). In Figures 3b and 3e, we demonstrate that these declines in word diversity are accompanied by similar declines in

combinatorial novelty; over time, the particular words that scientists and inventors use in the titles of their papers and patents are increasingly likely to have been used together in the titles of prior work. Consistent with these trends in language, we also observe declining novelty in the combinations of prior work cited by papers and patents, based on a previously established measure of “atypical combinations”<sup>32</sup> (Extended Data Figure 4).

The decline in disruptive activity is also apparent in the specific words used by scientists and inventors. If disruptiveness is declining, we reasoned that verbs alluding to the creation, discovery, or perception of new things should be used less frequently over time, whereas verbs alluding to the improvement, application, or assessment of existing things may be used more often<sup>30,31</sup>. Figure 3 shows the most common verbs in paper (Figure 3c) and patent titles (Figure 3f) in the first and last decade of each sample (see also Supplementary Information Section 2). While precisely and quantitatively characterizing words as “consolidating” or “disruptive” is challenging in the absence of context, the figure highlights a clear and qualitative shift in language. In the earlier decades, verbs evoking creation (e.g., “produce”, “form”, “prepare”, “make”), discovery (e.g., “determine”, “report”), and perception (e.g., “measure”) are prevalent in both paper and patent titles. In the later decades, however, these verbs are almost completely displaced by those tending to be more evocative of the improvement (e.g., “improve”, “enhance”, “increase”), application (e.g., “use”, “include”), or assessment (e.g., “associate”, “mediate”, “relate”) of existing scientific and technological knowledge and artifacts. Taken together, these patterns suggest a substantive shift in science and technology over time, with discovery and invention becoming less disruptive in nature, consistent with our results using the CD index.

## Conservation of highly disruptive work

The aggregate trends we document mask considerable heterogeneity in the disruptiveness of individual papers and patents and remarkable stability in the absolute number of highly disruptive works (see Methods, “Highly disruptive papers and patents,” and Figure 4). Specifically, despite large increases in scientific productivity, the number of papers and patents with  $CD_5$  values in the far right tail of the distribution remains nearly constant over time. This “conservation” of the absolute number of highly disruptive papers and patents holds despite considerable churn in the underlying fields responsible for producing those works (see Extended Data Figure 5, inset). These results suggest that the persistence of major breakthroughs—e.g., measurement of gravity waves, COVID-19 vaccines—is not inconsistent with slowing innovative activity. In short, declining aggregate disruptiveness does not preclude individually highly disruptive works.

## Alternative explanations

What is driving the decline in disruptiveness? Earlier, we suggested our results are not consistent with explanations that link slowing innovative activity to diminishing “low-hanging fruit.” Extended Data Figure 5 further shows that the decline in disruptiveness is unlikely due to other field-specific factors by decomposing variation in  $CD_5$  attributable to field, author, and year effects (see Methods, “Relative contribution of field, year, and author/inventor effects”).

Declining rates of disruptive activity are unlikely caused by the diminishing quality of science and technology<sup>17,33</sup>. If they were, then the patterns seen in Figure 2 should be less visible in high quality work. However, when we restrict our sample to articles published in premier publication venues like *Nature*, *Proceedings of the National Academy of Sciences*, and *Science* or to Nobel-winning discoveries (Figure 5)<sup>34</sup>, the downward trend persists.

Furthermore, the trend is not driven by characteristics of the WoS and UPSTO data or our particular derivation of the CD index; we observe similar declines in disruptiveness when we compute  $CD_5$  on papers in JSTOR, the American Physical Society corpus, Microsoft Academic Graph, and PubMed (see Methods, “Alternative samples”), the results of which are shown in Extended Data Figure 6. We further show that the decline is not an artifact of the CD index by reporting similar patterns using alternative derivations<sup>35,36</sup> (see Methods, “Alternative bibliometric measures,” and Extended Data Figure 7).

Declines in disruptiveness are also not attributable to changing publication, citation, or authorship practices (see Methods, “Robustness to changes in publication, citation, and authorship practices”). First, using approaches from the bibliometrics literature<sup>37–41</sup>, we computed several normalized versions of the CD index that adjusted for the increasing tendency for papers and patents to cite prior work<sup>42,43</sup>. Results using these alternative indicators (see Extended Data Figure 8a and 8d) were similar to those we reported in Figure 2. Second, using regression, we estimated models of  $CD_5$  as a function of indicator variables for each paper or patent’s publication year, along with specific controls for field  $\times$  year level—number of new papers/patents, mean number of papers/patents cited, mean number of authors/inventors per paper—and paper/patent-level—number of papers/patents cited—factors. Predictions from these models indicated a decline in disruptive papers and patents (see Extended Data Figure 8b and 8e and Supplementary Table 1) that was consistent with our main results. Finally, using Monte Carlo simulations, we randomly rewired the observed citation networks while preserving key characteristics of scientists’ and inventors’ citation behavior, including the number of citations made and received by individual papers and patents and the age gap between citing and cited works. We find that observed  $CD_5$  values are lower than those from the simulated networks (Extended Data Figure 8c and 8f), and the gap is widening: over time, papers and patents are increasingly less disruptive than would be expected by chance. Taken together, these additional analyses indicate that the decline in  $CD_5$  is unlikely to be driven by changing publication, citation, or authorship practices.

## Growth of knowledge and disruptiveness

We also considered how declining disruptiveness relates to the growth of knowledge (Extended Data Figure 9). On the one hand, scientists and inventors face an increasing knowledge burden, which may inhibit discoveries and inventions that disrupt the status quo. On the other hand, as previously noted, philosophers of science suggest that existing knowledge fosters discovery and invention<sup>3–5</sup>. Using regression models, we evaluated the relationship between the stock of papers and patents (a proxy for knowledge) within fields and their  $CD_5$  (Supplementary Information Section 3 and Supplementary Table 2). Interestingly, we find a positive effect of the growth of knowledge on disruptiveness for papers, consistent with prior work<sup>15</sup>; however, we find a negative effect for patents.

Given these conflicting results, we considered the possibility that the availability of knowledge may differ from its use. In particular, the growth in publishing and patenting may lead scientists and inventors to focus on narrower slices of prior work<sup>13,44</sup>, thereby limiting the “effective” stock of knowledge. Using three proxies, we document a decline in the use of prior knowledge among scientists and inventors (Figure 6). First, we see a decline in the diversity of work cited (Figure 6a and d), indicating that contemporary science and technology are engaging with narrower slices of existing knowledge. Moreover, this decline in diversity is accompanied by an increase in the share of citations to the 1% most highly cited papers and patents (Figures 6a1 and 6d1), which are also decreasing in semantic diversity (Figures 6a2 and 6d2). Over time, scientists and inventors are increasingly citing the same prior work, and that prior work is becoming more topically similar. Second, we see an increase in self-citation (Figure 6b and e), a common proxy for the continuation of one’s pre-existing research stream<sup>45–47</sup>, which is consistent with scientists and inventors relying more on highly familiar knowledge. Third, the mean age of work cited, a common measure for the use of dated knowledge<sup>48–50</sup>, is increasing (Figure 6c and f) suggesting that scientists and inventors may be struggling to keep up with the pace of knowledge expansion and instead relying on older, familiar work. All three indicators point to a consistent story: a narrower scope of existing knowledge is informing contemporary discovery and invention.

Results from a subsequent series of regression models suggest that use of less diverse work, more of one’s own work, and older work are all negatively associated with disruption (see Methods, “Regression analysis,” and Extended Data Table 1, and Supplementary Table 3), a pattern that holds even after accounting for the average age and number of prior works produced by team members. When the range of work used by scientists and inventors narrows, disruptive activity declines.

## Discussion

In summary, we report a dramatic decline in disruptive science and technology over time. Our analyses show that this trend is unlikely to be driven by changes in citation practices or the quality of published work. Rather, the decline

represents a substantive shift in science and technology, one that reinforces concerns about slowing innovative activity. We attribute this trend in part to scientists' and inventors' reliance on a narrower set of existing knowledge. While philosophers of science may be correct that the growth of knowledge is an endogenous process—wherein accumulated understanding promotes future discovery and invention—engagement with a broad range of extant knowledge is necessary for that process to play out, a requirement that appears more difficult with time. Relying on narrower slices of knowledge benefits individual careers<sup>51</sup>, but not scientific progress more generally.

Moreover, while the prevalence of disruptive works has declined, we find that the sheer number has remained stable. On the one hand, this result may suggest that there is a fixed “carrying capacity” for highly disruptive science and technology, in which case, policy interventions aimed at increasing such work may prove challenging. On the other hand, our observation of considerable churn in the underlying fields responsible for producing disruptive science and technology suggest the potential importance of factors like the shifting interests of funders and scientists and the “ripeness” of scientific and technological knowledge for breakthroughs, in which case the production of disruptive work may be responsive to policy levers. In either case, the stability we observe in the sheer number of disruptive papers and patents suggest that science and technology do not appear to have reached the end of the “endless frontier.” Room remains for the regular rerouting that disruptive works contribute to scientific and technological progress.

Our study is not without limitations. Notably, while research to date supports the validity of the CD index<sup>12,29</sup>, it is a relatively new indicator of innovative activity and will benefit from future work on its behavior and properties, especially across data sources and contexts. Studies that systematically examine the effect of different citation practices<sup>52,53</sup>, which vary across fields, would be particularly informative.

Overall, our results deepen understanding of the evolution of knowledge and may guide career planning and science policy. To promote disruptive science and technology, scholars may be encouraged to read widely and given time to keep up with the rapidly expanding knowledge frontier. Universities may forgo the focus on quantity, and more strongly reward research quality<sup>54</sup>, and perhaps more fully subsidize year-long sabbaticals. Federal agencies may invest in the riskier and longer-term individual awards that support careers and not simply specific projects<sup>55</sup>, giving scholars the gift of time needed to step outside the fray, inoculate themselves from the publish or perish culture, and produce truly consequential work. Understanding the decline in disruptive science and technology more fully permits a much-needed rethinking of strategies for organizing the production of science and technology in the future.

## References

1. Fleming, L. Recombinant uncertainty in technological search. *Management Science* **47**, 117–132 (2001).
2. Schumpeter, J. *Capitalism, Socialism and Democracy* (Perennial, 1942).
3. Koyré, A. An Unpublished Letter of Robert Hooke to Isaac Newton. *Isis* **43**, 312–337 (1952).
4. Weitzman, M. L. Recombinant growth. *Quarterly Journal of Economics* **113**, 331–360 (1998).
5. Acemoglu, D., Akcigit, U. & Kerr, W. R. Innovation network. *Proceedings of the National Academy of Sciences* **113**, 11483–11488 (2016).
6. Popper, K. *Conjectures and refutations: The growth of scientific knowledge* (Routledge, 2014).
7. Fleck, L. *Genesis and development of a scientific fact* (University of Chicago Press, 2012).
8. Tria, F., Loreto, V., Servedio, V. D. P. & Strogatz, S. H. The dynamics of correlated novelties. *Scientific Reports* **4**, 1–8 (2014).
9. Fink, T., Reeves, M., Palma, R. & Farr, R. Serendipity and strategy in rapid innovation. *Nature Communications* **8**, 1–9 (2017).
10. Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nature Reviews Drug Discovery* **10**, 428–438 (2011).
11. Bloom, N., Jones, C. I., Van Reenen, J. & Webb, M. Are ideas getting harder to find? *American Economic Review* **110**, 1104–44 (2020).
12. Funk, R. J. & Owen-Smith, J. A dynamic network measure of technological change. *Management Science* **63**, 791–817 (2017).
13. Jones, B. F. The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder? *Review of Economic Studies* **76**, 283–317 (2009).
14. Gordon, R. J. *The rise and fall of American growth* (Princeton University Press, 2016).

15. Chu, J. S. G. & Evans, J. A. Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences* **118** (2021).
16. Packalen, M. & Bhattacharya, J. NIH funding and the pursuit of edge science. *Proceedings of the National Academy of Sciences* **117**, 12011–12016 (2020).
17. Jaffe, A. B. & Lerner, J. *Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it* (Princeton University Press, 2011).
18. Horgan, J. *The end of science: Facing the limits of knowledge in the twilight of the scientific age* (Basic Books, 2015).
19. Collison, P. & Nielsen, M. Science is getting less bang for its buck. *Atlantic* (2018).
20. OECD. Effective policies to foster High-risk/High-reward research. *OECD Science, Technology, and Industry Policy Papers* (2021).
21. Nolan, A. *Artificial intelligence and the future of science* (ed OECD.AI) <https://oecd.ai/en/wonk/ai-future-of-science>. Oct. 2021.
22. Cowen, T. *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better* (Penguin, 2011).
23. Einstein, A. *The world as I see it* (Citadel Press, 1949).
24. Anderson, P. & Tushman, M. L. Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative Science Quarterly*, 604–633 (1990).
25. Arthur, W. B. The structure of invention. *Research Policy* **36**, 274–287 (2007).
26. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **140**, A1133 (1965).
27. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

28. Bornmann, L. & Tekles, A. Disruption index depends on length of citation window. *El profesional de la información (EPI)* **28** (2019).
29. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
30. Kuhn, T. S. *The Structure of Scientific Revolutions* (University of Chicago Press, 1962).
31. Brad Wray, K. Kuhn and the discovery of paradigms. *Philosophy of the Social Sciences* **41**, 380–397 (2011).
32. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
33. Ioannidis, J. P. Why most published research findings are false. *PLoS Medicine* **2**, e124 (2005).
34. Li, J., Yin, Y., Fortunato, S. & Wang, D. A dataset of publication records for Nobel laureates. *Scientific Data* **6**, 1–10 (2019).
35. Bornmann, L., Devarakonda, S., Tekles, A. & Chacko, G. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quantitative Science Studies* **1**, 1242–1259 (2020).
36. Leydesdorff, L., Tekles, A. & Bornmann, L. A proposal to revise the disruption index. *Profesional de la información (EPI)* **30** (2021).
37. Bornmann, L. & Marx, W. Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? *Journal of Informetrics* **9**, 408–418 (2015).
38. Waltman, L. & van Eck, N. J. in *Springer handbook of science and technology indicators* 281–300 (Springer, 2019).
39. Waltman, L. A review of the literature on citation impact indicators. *Journal of informetrics* **10**, 365–391 (2016).

40. Bornmann, L. How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles. *Quantitative Science Studies* **1**, 1553–1569 (2020).
41. Petersen, A. M., Pan, R. K., Pammolli, F. & Fortunato, S. Methods to account for citation inflation in research evaluation. *Research Policy* **48**, 1855–1865 (2019).
42. Bornmann, L. & Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology* **66**, 2215–2222 (2015).
43. Bornmann, L., Haunschild, R. & Mutz, R. Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications* **8**, 1–15 (2021).
44. Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *Proceedings of the national academy of sciences* **108**, 18910–18914 (2011).
45. Bonzi, S. & Snyder, H. Motivations for citation: A comparison of self citation and citation to others. *Scientometrics* **21**, 245–254 (1991).
46. Fowler, J. & Aksnes, D. Does self-citation pay? *Scientometrics* **72**, 427–437 (2007).
47. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. & West, J. D. Men set their own cites high: Gender and self-citation across fields and over time. *Socius* **3**, 2378023117738903 (2017).
48. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow’s breakthroughs in science and technology: The hotspot. *Science Advances* **3**, e1601315 (2017).
49. Merton, R. K. Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society* **105**, 470–486 (1961).

50. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
51. Leahey, E. Not by productivity alone: How visibility and specialization contribute to academic earnings. *American sociological review* **72**, 533–561 (2007).
52. Tahamtan, I. & Bornmann, L. Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of informetrics* **12**, 203–216 (2018).
53. Tahamtan, I. & Bornmann, L. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics* **121**, 1635–1684 (2019).
54. Bhattacharya, J. & Packalen, M. *Stagnation and scientific incentives* tech. rep. (National Bureau of Economic Research, 2020).
55. Azoulay, P., Graff Zivin, J. S. & Manso, G. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics* **42**, 527–554 (2011).

## Captions

**Figure 1: Overview of the measurement approach.** This figure presents a schematic visualization of the CD index. **a** shows the CD index value of three Nobel-prize-winning papers and three notable patents in our sample, measured as of 5 years post-publication (indicated by  $CD_5$ ). **b** shows the distribution of  $CD_5$  for papers from WoS ( $n=24,659,076$ ) between 1945-2010 and patents from Patents View ( $n=3,912,353$ ) between 1976-2010, where a single dot represents a paper or patent. The vertical (up/down) dimension of each “strip” corresponds to values of the CD index (with axis values shown in orange on the left). The horizontal (left/right) dimension of each strip is included to minimize overlapping of the points. Darker areas on each strip plot indicate denser regions of the distribution (i.e., more commonly observed  $CD_5$  values). Additional details on the distribution of the CD index are given in Extended Data Figure 1. **c** shows three hypothetical citation networks, where the CD index is at its the maximally disruptive value ( $CD_t = 1$ ), midpoint value ( $CD_t = 0$ ), and maximally consolidating value ( $CD_t = -1$ ). The panel also provides the equation for the CD index and an illustrative calculation.

**Figure 2: Decline of disruptive science and technology.** This figure shows the decline in  $CD_5$  over time, separately for papers (**a**,  $n=24,659,076$ ) and patents (**b**,  $n=3,912,353$ ). For papers, lines correspond to WoS research areas; from 1945-2010 the magnitude of decline ranges from 91.9% (Social Science) to 100% (Physical Science). For patents, lines correspond to National Bureau of Economic Research (NBER) technology categories; from 1980-2010 the magnitude of decline ranges from 93.5% (Computers and Communications) to 96.4% (Drugs and Medical). Shaded bands correspond to 95% confidence intervals. As we elaborate in the Methods (“Robustness to changes in publication, citation, and authorship practices”), this pattern of decline is robust to adjustment for confounding from changes in publication, citation, and authorship practices over time.

**Figure 3: The decline of disruptive science and technology is visible in the changing language of papers and patents.** **a** and **d** show a decline in the diversity of language used in science and technology based on the unique/total words of paper titles from 1945-2010 (**a**,  $n=24,659,076$ ) and of patent titles from 1980-2010 (**d**,  $n=3,912,353$ ). **b** and **e** show a decline in the novelty of language used in science and technology based on the number of new word pairs/total word pairs introduced each year in WoS paper titles from 1945-2010 (**b**) and in Patents View patent titles from 1980-2010 (**e**). For papers in both **a** and **b**, lines correspond to WoS research areas ( $n=264$  WoS research area  $\times$  year observations). For patents in both **d** and **e**, lines correspond to NBER technology categories ( $n=229$  NBER technology category  $\times$  year observations). **c** and **f** show the frequency of the most commonly used verbs in paper titles for the first (red) and last (blue) decades of the observation period in paper (**c**,  $n=24,659,076$ ) and patent (**f**,  $n=3,912,353$ ) titles.

**Figure 4: Conservation of highly disruptive work.** This figure shows the number of disruptive papers (**a**,  $n=5,030,179$ ) and patents (**b**,  $n=1,476,004$ ) across four different ranges of  $CD_5$  (papers and patents with  $CD_5$  values in the range  $[-1.0, 0.0)$  are not represented in the figure). Lines correspond to different levels of disruptiveness as measured by  $CD_5$ . Despite substantial increases in the numbers of papers and patents published each year, there is little change in the number of highly disruptive papers and patents, as evidenced by the relatively flat red, green, and orange lines. This pattern helps to account for simultaneous observations of both aggregate evidence of slowing

innovative activity and seemingly major breakthroughs in many fields of science and technology. The inset plots show the composition of the most disruptive papers and patents (defined as those with  $CD_5$  values  $> 0.25$ ) by field over time. The observed stability in the absolute number of highly disruptive papers and patents holds despite considerable churn in the underlying fields of science and technology responsible for producing those works.

**Figure 5: CD index of high-quality science over time.** This figure shows changes in  $CD_5$  over time for papers published in *Nature*, *Proceedings of the National Academy of Sciences* (PNAS), and *Science* (inset plot,  $n=223,745$ ) and Nobel-Prize-winning papers (main plot,  $n=635$ ). Colors indicate the three different journals in the inset plot; colors indicate the three different fields in which the Nobel Prize is awarded in the main plot. Shaded bands correspond to 95% confidence intervals. For historical completeness, we plot CD index scores for all Nobel papers back to 1900 (the first year in which the prize was awarded); however, our main analyses begin in the post-1945 era, when the Web of Science data are generally more reliable. The figure indicates that changes in the quality of published science over time is unlikely to be responsible for the decline in disruption.

**Figure 6: Papers and patents are using narrower portions of existing knowledge.** This figure shows changes in the level of diversity of existing scientific and technological knowledge use among papers (**a**,  $n=264$  WoS research area  $\times$  year observations; **b** and **c**,  $n=24,659,076$  papers) and patents (**d**, 229 NBER technology category  $\times$  year observations; **e** and **f**,  $n=3,912,353$  patents) based on following measures: diversity of work cited (**a** and **d**), mean number of self-citations (**b** and **e**), and mean age of cited work (**c** and **f**). Shaded bands (panels **b**, **c**, **e**, and **f**) correspond to 95% confidence intervals. The inset plots of **a** and **d** show changes in the share of citations to the top 1% most highly cited papers (**a1** and **d1**) and in the semantic diversity of the top 1% most cited over time (**a2** and **d2**). Values of both measures are computed within field and year, and are subsequently averaged across fields for plotting. Semantic diversity is based on paper and patent titles; values correspond to the ratio of the standard deviation to the mean pairwise cosine similarity (i.e., the coefficient of variation) among the titles of the 1% most cited papers and patents by field and year. To enable semantic comparisons, titles were vectorized using pretrained word embeddings. For papers, lines are shown for each WoS research area; for patents, lines are shown for each NBER technology category. In subsequent regression analyses using these measures, we find that using less diverse work, more of one’s own work, and older work is associated with less disruptive papers and patents (see Methods, “Regression analysis,” and Extended Data Table 1).

**Extended Data Figure 1: Distribution of  $CD_5$ .** This figure gives an overview of the distribution of  $CD_5$  for papers ( $n=24,659,076$ ) and patents ( $n=3,912,353$ ). Panels **a** and **c** show counts of papers and patents over discrete intervals of  $CD_5$ . Panels **b** and **d** show the distribution of  $CD_5$  over time, within 10 (papers) and 5 (patents) year intervals, using letter-value plots. These plots are similar to boxplots, but generally provide more reliable summaries for large datasets. They are drawn by identifying the median of the underlying distribution and then recursively drawing boxes outward from there in either direction that encompass half of the remaining data.

**Extended Data Figure 2: CD index measured using alternative forward citation windows.** This figure evaluates the sensitivity of our results to the use of different forward citation windows when computing the CD index for papers ( $n=24,659,076$ ) and patents ( $n=3,912,353$ ). In the main text, the index is computed based on citations made to papers and patents and their backward

references as of 5 years after the year of publication. **a** and **c** plot the CD index using a longer, 10 year forward window, for papers and patents, respectively. **b** and **d** plot the CD index using all forward citations made to sample papers and patents as of the year 2017. Shaded bands correspond to 95% confidence intervals. Overall, the results mirror those reported in the main text, although the decline is somewhat steeper using longer forward citation windows, suggesting our primary results may represent a more conservative estimate.

**Extended Data Figure 3: Diversity of language use in science and technology over time.** This figure shows changes in the ratio of unique to total words (also known as the type-token ratio) over time based on data from the abstracts of papers (**a**,  $n=76$  WoS research area  $\times$  year observations) and patents (**b**,  $n=229$  NBER technology category  $\times$  year observations). For papers, lines correspond to WoS research areas; for patents, lines correspond to NBER technology categories. For paper abstracts, lines begin in 1992 because WoS does not reliably record abstracts for papers published prior to the early 1990s. The ratio of unique to total words is computed separately by field (i.e., the uniqueness of words and total word counts are determined within WoS research areas and NBER technology categories). If disruption is decreasing, we may plausibly expect to see a decrease in the diversity of words used by scientists and inventors, as discoveries and inventions will be less likely to create departures from the status quo, and will therefore be less likely to need to introduce new terminology. For both papers and patents, we observe declining diversity in word use over time, which is consistent with this expectation and corroborates our findings using the CD index.

**Extended Data Figure 4: Declining combinatorial novelty.** This figure shows changing patterns in the combinatorial novelty/conventionality of papers (**a**,  $n=24,659,076$ ) and patents (**b**,  $n=3,912,353$ ), using a previously proposed measure of “atypical combinations”<sup>32</sup>. The measure quantifies the degree to which the prior work cited by a paper or patent would be expected by chance. For papers, we follow prior work<sup>32</sup> and consider combinations of cited journals. For example, if a paper made three citations to prior work, and that work was published in three different journals—*Nature*, *Cell*, and *Science*—then there are three combinations—*Nature*  $\times$  *Cell*, *Nature*  $\times$  *Science*, and *Science*  $\times$  *Cell*. To determine the degree to which each combination would be expected by chance, the frequency of observed pairings are compared to those in 10 “rewired” copies of the overall citation network, using a z-score. For patents, there is no natural analogue to journals, and therefore we consider pairings of primary United States Patent Classification (USPC) system codes. We present the results of this analysis following the approach of prior work<sup>32</sup>, which plots the cumulative distribution function of the measure. In general, there is a rightward shift in the cumulative distributions over time, suggesting that for both papers and patents, combinations are more conventional than would be expected by chance, consistent with what we would anticipate based on our results using the CD index. For patents, there is also a smaller shift in the opposite direction on the left side of the distribution, suggesting that novel patents in recent decades are somewhat more novel than novel patents in earlier decades. Overall, however, the bulk of the distribution is moving to the right, indicating greater conventionality.

**Extended Data Figure 5: Contribution of field, year, and author effects.** This figure shows the relative contribution of field, year, and author fixed effects to the adjusted  $R^2$  in regression models predicting  $CD_5$ . The top bar shows the results for papers ( $n=80,607,091$  paper  $\times$  author observations); the bottom bar shows the results for patents ( $n=8,319,826$  patent  $\times$  inventor observations).

The results suggest that for both papers and patents, stable characteristics of authors contribute significantly to patterns of disruptiveness. Moreover, relatively little of the variation is accounted for by field-specific factors.

**Extended Data Figure 6: CD index over time across data sources.** This figure shows changes in  $CD_5$  over time across four additional data sources (the WoS [n=24,659,076] and Patents View [n=3,912,353] lines are included for reference): JSTOR (n=1,703,353), the American Physical Society corpus (n=478,373), Microsoft Academic Graph (n=1,000,000), and PubMedD (n=16,774,282). Colors indicate the six different data sources. Shaded bands correspond to 95% confidence intervals. The figure indicates that the decline in disruption is unlikely driven by our sample choice of WoS papers and Patents View patents.

**Extended Data Figure 7: Alternative measures of disruption.** This figure shows the decline in the disruption of papers (**a**, n=100,000) and patents (**b**, n=100,000) based on two alternative measures of disruption. The blue lines calculate disruption using a measure proposed in Bornmann et al.<sup>35</sup>,  $DI_l^{nok}$  where  $l = 5$ , which makes the measure more resilient to marginal changes in the number of papers or patents that only cite the focal work's references. The orange lines calculate disruption using a measure proposed in Leydesdorff et al.<sup>36</sup>,  $DI^*$ , which makes the measure less sensitive to small changes in the forward citation patterns of papers or patents that make no backward citations. Shaded bands correspond to 95% confidence intervals. With both alternative measures, we observe decreases in disruption for papers and patents, suggesting that the decline is not an artifact of our operationalization of disruption.

**Extended Data Figure 8: Robustness to changes in publication, citation, and authorship practices.** This figure evaluates whether the observed declines in disruptiveness may be attributable to changes in publication, citation, and authorship practices for papers (n=24,659,076) and patents (n=3,912,353). Panels **a** and **d** adjust for these changes using a normalization approach. We present two alternative versions of the CD index, both of which account for the tendency for papers and patents to cite more prior work over time. Blue lines indicate normalization at the paper level (accounting for the number of citations made by the focal paper or patent). Orange lines indicate normalization at the field and year level (accounting for the mean number of citations made by papers or patents in the focal field and year). Panels **b** (papers) and **e** (patents) adjust for changes in publication, citation, and authorship practices using a regression approach. The panels show predicted values of  $CD_5$  based on regressions reported in Models 4 (papers) and 8 (patents) of Supplementary Table 1, which adjust for field  $\times$  year—*Number of new papers/patents*, *Mean number of papers/patents cited*, *Mean number of authors/inventors per paper/patent*—and paper/patent-level—*Number of papers/patents cited*, *Number of unlinked references*—characteristics. Predictions are made separately for each of the year indicators included in the models; we then connect these separate predictions with lines to aid interpretation. Finally, Panels **c** (papers) and **f** (patents) adjust for changes in publication, citation, and authorship practices using a simulation approach. The panels plot z-scores that compare values of  $CD_5$  obtained from the observed citation networks to those obtained from randomly rewired copies of the observed networks. Across all six panels, shaded bands correspond to 95% confidence intervals.

**Extended Data Figure 9: Growth of scientific and technological knowledge.** This figure shows the number of papers (n=24,659,076) published (**a**) and patents (n=3,912,353) granted (**b**) over

time. For papers, lines correspond to WoS research areas; for patents, lines correspond to NBER technology categories.

## Methods

**WoS data** We limit our focus to research papers published between 1945 and 2010. Although the WoS data begins in the year 1900, the scale and social organization of science shifted dramatically in the post-war era, thereby making comparisons with the present difficult and potentially misleading<sup>56–58</sup>. We end our analyses of papers in 2010 because some of our measures require several subsequent years of data following paper publication. The WoS data archives 65 million documents published in 28,968 journals between 1900 to 2017 and 735 million citations among them. In addition, the WoS data include the titles and the full text of abstracts for 65 and 29 million records, respectively, published between 1913 to 2017. After eliminating non-research documents (e.g., book reviews, commentaries) and subsetting the data to the 1945-2010 window, the analytical sample consists of  $n=24,659,076$  papers.

**Patents View data** We limit our focus to patents granted from 1976, which is the earliest year for which machine-readable records are available in the Patents View data. As we did with papers, we end our analyses in 2010 because some measures require data from subsequent years for calculation. The Patents View data is the most exhaustive source of historical data on inventions, with information on 6.5 million patents granted between 1976 and 2017 and their corresponding 92 million citations. The Patents View data includes the titles and abstracts for 6.5 million patents granted between 1976 and 2017. Following prior work<sup>12</sup>, we focused our attention on utility patents, which cover the vast majority (91% in our data) of patented inventions. After eliminating non-utility patents and subsetting the data to the 1976-2010 window, the analytical sample consists of  $n=3,912,353$  patents.

**Highly disruptive papers and patents** Observations (and claims) of slowing progress in science and technology are increasingly common, supported not just by the evidence we report, but also by prior research from diverse methodological and disciplinary perspectives<sup>10,11,13–19</sup>. Yet as noted in the main text, there is a tension between observations of slowing progress from aggregate data on the one hand, and continuing reports of seemingly major breakthroughs in many fields of science and technology—spanning everything from the measurement of gravity waves to the sequencing of the human genome—on the other. In an effort to reconcile this tension, we considered the possibility that while overall, discovery and invention may be less disruptive over time, the high-level view taken in prior work may mask considerable heterogeneity. Put differently, aggregate evidence of slowing progress does not preclude the possibility that some (smaller) subset of discoveries and inventions are highly disruptive.

To evaluate this possibility, we plot the number of disruptive papers (Figure 4a) and patents (Figure 4b) over time, where disruptive papers and patents are defined as those with  $CD_5$  values  $> 0$ . Within each panel, we plot four lines, corresponding to four evenly spaced intervals—(0, 0.25], (0.25, 0.5], (0.5, 0.75], (0.75, 1.00]—over the positive values of  $CD_5$ . The first two intervals therefore correspond to papers and patents that are relatively weakly disruptive, while the latter two correspond to those that are more strongly so (e.g., where we may expect to see major breakthroughs like some of those mentioned above). Despite major increases in the numbers of papers and patents published each year, we see little change in the number of highly disruptive papers and patents, as evidenced by the relatively flat red, green, and orange lines. Remarkably,

this “conservation” of disruptive work holds even despite fluctuations over time in the composition of the scientific and technological fields responsible for producing the most disruptive work (see Figure 4, inset plots). Overall, these results help to account for simultaneous observations of both major breakthroughs in many fields of science and technology and aggregate evidence of slowing progress.

**Relative contribution of field, year, and author/inventor effects** Our results show a steady decline in the disruptiveness of science and technology over time. Moreover, the patterns we observe are generally similar across broad fields of study, which suggests that the factors driving the decline are not unique to specific domains of science and technology. The decline could be driven by other factors, such as the conditions of science and technology at a point in time or the particular individuals that produce science and technology. For example, exogenous factors like economic conditions may encourage research or invention practices that are less disruptive. Similarly, scientists and inventors of different generations may have different approaches, which may result in greater or lesser tendencies for producing disruptive work. We therefore sought to understand the relative contribution of field, year, and author (or inventor) factors to the decline of disruptive in science and technology.

To do so, we decomposed the relative contribution of field, year, and author fixed effects to the predictive power of regression models of the CD index. The unit of observation in these regressions is the author (or inventor)  $\times$  year. We enter field fixed effects using granular subfield indicators (i.e., 150 WoS subject areas for papers, 138 NBER subcategories for patents). For simplicity, we did not include additional covariates beyond the fixed effects in our models. Field fixed effects capture all field-specific factors that do not vary by author or year (e.g., the basic subject matter); year fixed effects capture all year-specific factors that do not vary by field or author (e.g., the state of communication technology); author (or inventor) fixed effects capture all author-specific factors that do not vary by field or year (e.g., the year of PhD awarding). After specifying our model, we determine the relative contribution of field, year, and author fixed effects to the overall model adjusted  $R^2$  using Shapley-Owen decomposition. Specifically, given our  $n = 3$  groups of fixed effects (field, year, and author) we evaluate the relative contribution of each set of fixed effects by estimating the adjusted  $R^2$  separately for the  $2^n$  models using subsets of the predictors. The relative contribution of each set of fixed effects is then computed using the Shapley value from game theory<sup>59</sup>.

Results of this analysis are shown in Extended Data Figure 5, for both papers (top bar) and patents (bottom bar). Total bar size corresponds to the value of the adjusted  $R^2$  for the fully specified model (i.e., with all three groups of fixed effects). Consistent with our observations from plots of the CD index over time, we observe that for both papers and patents, field specific factors make the lowest relative contribution to the adjusted  $R^2$  (0.02 and 0.01 for papers and patents, respectively). Author fixed effects, by contrast, appear to contribute much more to the predictive power of the model, for both papers (0.20) and patents (0.17). Researchers and inventors who enter the field in more recent years may face a higher burden of knowledge and thus resort to building on narrower slices of existing work (e.g., due to more specialized doctoral training), which would generally lead to less disruptive science and technology being produced in later years, consistent with our findings. The pattern is more complex for year fixed effects; while year-specific factors that do not vary by field or author hold more explanatory power than field for both papers (0.02) and

patents (0.16), they appear to be substantially more important for the latter than the former. Taken together, these findings suggest that relatively stable factors that vary across individual scientists and inventors may be particularly important for understanding changes in disruptiveness over time. The results also confirm that domain-specific factors across fields of science and technology play a very small role in explaining the decline in disruptiveness of papers and patents.

**Alternative samples** We also considered whether the patterns we document may be artifacts of our choice of data sources. While we observe consistent trends in both the WoS and Patents View data, and both databases are widely used by the Science of Science community, our results may conceivably be driven by factors like changes in coverage (e.g., journals added or excluded from WoS over time) or even data errors rather than fundamental changes in science and technology. To evaluate this possibility, we therefore calculated  $CD_5$  for papers in four additional databases—JSTOR, the American Physical Society corpus, Microsoft Academic Graph, and PubMed. We included all records from 1930–2010 from PubMed (16,774,282 papers), JSTOR (1,703,353 papers), and American Physical Society (478,373 papers). The JSTOR data were obtained via a special request from ITHAKA, the data maintainer (<http://www.ithaka.org>), as were the American Physical Society data (see <https://journals.aps.org/datasets>). We downloaded the Microsoft Academic Graph data from CADRE at Indiana University (<https://cadre.iu.edu/>). The PubMed data were downloaded from the National Library of Medicine FTP server (<ftp://ftp.ncbi.nlm.nih.gov/pubmed/baseline>). Due to the exceptionally large scale of Microsoft Academic Graph and the associated computational burden, we randomly extracted 1 million papers. As shown in Extended Data Figure 6, the downward trend in disruptiveness is evident across all samples.

**Alternative bibliometric measures** Several recent papers have introduced alternative specifications of CD index<sup>12</sup>. We evaluated whether the declines in disruptiveness we observe are corroborated using two alternative variations. One criticism of the CD index has been that the number of papers that cite only the focal paper’s references dominates the measure<sup>35</sup>. Bornmann et al.<sup>35</sup> proposes  $DI_l^{nok}$  as a variant that is less susceptible to this issue. Another potential weakness of the CD index is that it could be very sensitive to small changes in the forward citation patterns of papers that make no backward citations<sup>36</sup>. Leydesdorff et al.<sup>36</sup> suggests  $DI^*$  as an alternate indicator of disruption that addresses this issue. Therefore, we calculated  $DI_l^{nok}$  where  $l = 5$  and  $DI^*$  for 100,000 randomly drawn papers and patents each from our analytic sample. Results are presented in Extended Data Figure 7a (papers) and 9b (patents). The blue lines indicate disruption based on Bornmann et al.<sup>35</sup> and the orange lines indicate disruption based on Leydesdorff et al.<sup>36</sup>. Across science and technology, the two alternative measures both show declines in disruption over time, similar to the patterns observed with the CD index. Taken together, these results suggest that the declines in disruption we document are not an artifact of our particular operationalization.

**Robustness to changes in publication, citation, and authorship practices** We also considered whether our results may be attributable to changes in publication, citation, or authorship practices, rather than by substantive shifts in discovery and invention. Perhaps most critically, as noted in the main text, there has been a dramatic expansion in publishing and patenting over the period of our study. This expansion has naturally increased the amount of prior work that is rele-

vant to current science and technology and therefore at risk of being cited, a pattern reflected in the dramatic increase in the average number of citations made by papers and patents (i.e., papers and patents are citing more prior work than in previous eras)<sup>42,43</sup>. Recall that the CD index quantifies the degree to which future work cites a focal work together with its predecessors (i.e., the references in the bibliography of the focal work). Greater citation of a focal work independently of its predecessors is taken to be evidence of a social process of disruption. As papers and patents cite more prior work, however, the probability of a focal work being cited independently of its predecessors may decline mechanically; the more citations a focal work makes, the more likely future work is to cite it together with one of its predecessors, even by chance. Consequently, increases in the number of papers and patents available for citing and in the average number of citations made by scientists and inventors may contribute to the declining values of the CD index. In short, given the dramatic changes in science and technology over our long study window, the CD index of papers and patents published in earlier periods may not be directly comparable to those of more recent vintage, which could in turn render our conclusions about the decline in disruptive science and technology suspect. We addressed these concerns using three distinctive but complementary approaches—normalization, regression adjustment, and simulation.

*Verification using normalization.* First, following common practice in bibliometric research<sup>37–41</sup>, we developed two normalized versions of the CD index, with the goal of facilitating comparisons across time. Among the various components of the CD index, we focused our attention on the count of papers or patents that only cite the focal work’s references (“ $N_k$ ”), as this term would seem most likely to scale with the increases in publishing and patenting and in the average number of citations made by papers and patents to prior work<sup>35</sup>. Larger values of  $N_k$  lead to smaller values of the CD index. Consequently, dramatic increases in  $N_k$  over time, particularly relative to other components of the measure, may lead to a downward bias, thereby inhibiting our ability to accurately compare disruptive science and technology in later years with earlier periods.

Our two normalized versions of the CD index aim to address this potential bias by attenuating the effect of increases in  $N_k$ . In the first version, which we call “Paper normalized,” we subtract from  $N_k$  the number of citations made by the focal paper or patent to prior work (“ $N_b$ ”). The intuition behind this adjustment is that when a focal paper or patent cites more prior work,  $N_k$  is likely to be larger because there are more opportunities for future work to cite the focal paper or patent’s predecessors. This increase in  $N_k$  would result in lower values of the CD index, though not necessarily as a result of the focal paper or patent being less disruptive. In the second version, which we call “Field  $\times$  year normalized,” we subtract  $N_k$  by the average number of backward citations made by papers or patents in the focal paper or patent’s WoS research area or NBER technology category, respectively, during its year of publication (we label this quantity “ $N_b^{mean}$ ”). The intuition behind this adjustment is that in fields and time periods in which there is a greater tendency for scientists and inventors to cite prior work,  $N_k$  is also likely to be larger, thereby leading to lower values of the CD index, though again not necessarily as a result of the focal paper or patent being less disruptive. In cases where either  $N_b$  or  $N_b^{mean}$  exceed the value of  $N_k$ , we set  $N_k$  to 0 (i.e.,  $N_k$  is never negative in the normalized measures). Both adaptations of the CD index are inspired by established approaches in the scientometrics literature, and may be understood as a form of “citing side normalization” (i.e., normalization by correcting for the effect of differences in lengths of references lists<sup>39</sup>).

In Extended Data Figure 8, we plot the average values of both normalized versions of the CD index over time, separately for papers (8a) and Patents (8d). Consistent with our findings reported

in the main text, we continue to observe a decline in the CD index over time, suggesting that the patterns we observe in disruptive science and technology are unlikely driven changes in citation practices.

*Verification using regression adjustment.* Second, we adjusted for potential confounding using a regression-based approach. This approach complements the bibliometric normalizations just described by allowing us to account for a broader array of changes in publication, citation, and authorship practices in general (the latter of which is not directly accounted for in either the normalization approach or the simulation approach described next), and increases the amount of prior work that is relevant to current science and technology in particular. In Supplementary Table 1, we report the results of regression models predicting  $CD_5$  for papers (Models 1-4) and patents (Models 5-8), with indicator variables included for each year of our study window (the reference categories are 1945 and 1980 for papers and patents, respectively). Models 1 and 4 are the baseline models, and include no other adjustments beyond the year indicators. In Models 2 and 5, we add subfield fixed effects (Web of Science subject areas for papers, NBER technology subcategories for patents). Finally, in Models 3-4 and 7-8, we add control variables for several field  $\times$  year level—*Number of new papers/patents*, *Mean number of papers/patents cited*, *Mean number of authors/inventors per paper*—and paper/patent-level—*Number of papers/patents cited*—characteristics, thereby enabling more robust comparisons in patterns of disruptive science and technology over the long time period spanned by our study. For the paper models, we also include a paper-level control for the *Number of unlinked references* (i.e., the number of citations to works that are not indexed in WoS). We find that the inclusion of these controls improves model fit, as indicated by statistically significant Wald tests presented below the relevant models.

Across all 8 models shown in Supplementary Table 1, we find that the coefficients on the year indicators are statistically significant and negative, and growing in magnitude over time, which is consistent with the patterns we reported based on unadjusted  $CD_5$  values index in the main text (i.e., Figure 2). In Extended Data Figure 8, we visualize the results of our regression-based approach by plotting the predicted  $CD_5$  values separately for each of the year indicators included in Models 4 (papers) and 8 (patents). To enable comparisons with raw  $CD_5$  values shown in the main text, we present the separate predictions made for each year as a line graph. As shown in the figure, we continue to observe declining values of the CD index across papers and patents, even when accounting for changes in publication, citation, and authorship practices.

*Verification using simulation.* Third, following related work in the Science of Science<sup>32,60–62</sup>, we considered whether our results may be an artifact of changing patterns in publishing and citation practices by using a simulation approach. In essence, the CD index measures disruption by characterizing the network of citations around a focal paper or patent. However, many complex networks, even those resulting from random processes, exhibit structures that yield nontrivial values on common network measures (e.g., clustering)<sup>63–65</sup>. During the period spanned by our study, the citation networks of science and technology experienced significant change, with dramatic increases in both the numbers of nodes (i.e., papers or patents) and edges (i.e., citations). Thus, rather than reflecting a meaningful social process, the observed declines in disruption may result from these structural changes in the underlying citation networks.

To evaluate this possibility, we followed standard techniques from network science<sup>63,66</sup> and conducted an analysis in which we recomputed the CD index on randomly rewired citation networks. If the patterns we observe in the CD index are the result of structural changes in the citation networks of science and technology (e.g., growth in the number of nodes or edges) rather than a

meaningful social process, then these patterns should also be visible in comparable random networks that experience similar structural changes. Therefore, finding that the patterns we see in the CD index differ for the observed and random citation networks would serve as evidence that the decline in disruption is not an artifact of the data.

We began by creating copies of the underlying citation network on which the values of the CD index used in all analyses reported in the main text were based, separately for papers and patents. For each citation network (one for papers, one for patents), we then rewired citations using a degree-preserving randomization algorithm. In each iteration of the algorithm, two edges (e.g., A-B, C-D) are selected from the underlying citation network, after which the algorithm attempts to swap the two endpoints of the edges (e.g., A-B becomes A-D, and C-D becomes C-B). If the degree centrality of A, B, C, and D remains the same after the swap, the swap is retained; otherwise, the algorithm discards the swap and moves on to the next iteration. When evaluating degree centrality, we consider “in-degree” (i.e., citations from other papers/patents to the focal paper/patent) and “out-degree” (i.e., citations from the focal paper/patent to other papers/patents) separately. Furthermore, we also required that the age distribution of citing and cited papers/patents was identical in the original and rewired networks. Specifically, swaps were only retained when the publication year of the original and candidate citations were the same. In light of these design choices, our rewiring algorithm should be seen as fairly conservative, as it preserves substantial structure from the original network. There is no scholarly consensus on the number of swaps necessary to ensure the original and rewired networks are sufficiently different from one another; the rule we adopt here is  $100 \times m$ , where  $m$  is the number of edges in the network being rewired.

Following previous work<sup>32</sup>, we created 10 rewired copies of the observed citation networks for both papers and patents. After creating these rewired citation networks, we then recomputed  $CD_5$ . Due to the large scale of the WoS data, we base our analyses on a random subsample of 10 million papers;  $CD_5$  was computed on the rewired network for all patents. For each paper and patent, we then compute a z-score that compares the observed  $CD_5$  value to those of the same paper or patent in the 10 rewired citation networks. Positive z-scores indicate that the observed  $CD_5$  value is greater (i.e., more disruptive) than would be expected by chance; negative z-scores indicate that the observed values are lesser (i.e., more consolidating).

The results of these analyses are shown in Extended Data Figure 8, separately for papers (Extended Data Figure 8c) and patents (Extended Data Figure 8f). Lines correspond to the average z-score among papers or patents published in the focal year. The plots reveal a pattern of change in the CD index over and beyond that “baked in” to the changing structure of the network. We find that on average, papers and patents tend to be less disruptive than would be expected by chance, and moreover, the gap between the observed CD index values and those from the randomly rewired networks is increasing over time, which is consistent with our findings of a decline in disruptive science and technology.

Taken together, the results of the foregoing analyses suggest that while there have been dramatic changes in science and technology over the course of our long study window, particularly with respect to publication, citation, and authorship practices, the decline in disruptive science and technology that we document using the CD index is unlikely an artifact of these changes, and instead represents a substantive shift in the nature of discovery and invention.

**Regression analysis** We evaluate the relationship between disruptiveness and the use of prior knowledge using regression models, predicting  $CD_5$  for individual papers and patents, based on three indicators of prior knowledge use—the diversity of work cited, mean number of self-citations, and mean age of work cited. Our measure of the diversity of work cited is measured at the field  $\times$  year level; all other variables included in the regressions are defined at the level of the paper or patent. To account for potential confounding factors, our models included year and field fixed effects. Year fixed effects account for time variant factors that affect all observations (papers or patents) equally (e.g., global economic trends). Field fixed effects account for field-specific factors that do not change over time (e.g., some fields may intrinsically value disruptive work over consolidating ones). In contrast to our descriptive plots, for our regression models, we adjust for field effects using the more granular 150 WoS “extended subjects” (e.g., “Biochemistry & Molecular Biology,” “Biophysics,” “Biotechnology & Applied Microbiology,” “Cell Biology, Developmental Biology,” “Evolutionary Biology,” and “Microbiology” are extended subjects within the “Life Sciences & Biomedicine” research area) and 38 NBER technology subcategories (e.g., “Agriculture, Food, Textile,” “Coating,” “Gas,” “Organic,” and “Resins” are subcategories within the “Chemistry” technology category).

In addition, we also include controls for the *Mean age of team members* (i.e., “career age,” defined as the difference between the publication year of the focal paper or patent and the first year in which each author or inventor published a paper or patent) and the *Mean number of prior works produced by team members*. While increases in rates of self citations may indicate that scientists and inventors are becoming more narrowly focused on their own work, these rates may also be driven in part by the amount of prior work available for self-citing. Similarly, while increases in the age of work cited in papers and patents may indicate that scientists and inventors are struggling to keep up, they may also be driven by the rapidly aging workforce in science and technology<sup>67,68</sup>. For example, older scientists and inventors may be more familiar with or more attentive to older work, or may actively resist change<sup>69</sup>. These control variables help to account for these alternative explanations.

Supplementary Table 3 shows summary statistics for variables used in the ordinary-least-squares regression models. The diversity of work cited is measured by normalized entropy, which ranges from 0 to 1. Greater values on this measure indicate a more uniform distribution of citations to a wider range of existing work; lower values indicate a more concentrated distribution of citations to a smaller range of existing work. The tables show that the normalized entropy in a given field and year has a nearly maximal average entropy of 0.98 for both science and technology. About 16% of papers cited in a paper are by an author of the focal paper; the corresponding number for patents is about 7%. Papers tend to rely on older work and work that varies more greatly in age (measured by standard deviation) than patents. Additionally, the average  $CD_5$  of a paper is 0.04 while the average  $CD_5$  of a patent is 0.12, meaning that the average paper tends to be less disruptive than the average patent.

We find that using more diverse work, less of one’s own work, and older work tends to be associated with the production of more disruptive science and technology, even after accounting for the average age and number of prior works produced by team members. These findings are based on our regression results, shown in Extended Data Table 1. Models 6 and 12 present the full regression models. The models indicate a consistent pattern for both science and technology, wherein the coefficients for diversity of work cited are positive and significant for papers (0.159,  $p < 0.01$ ) and patents (0.069,  $p < 0.01$ ), indicating that in fields where there is more use of diverse

work, there is greater disruption. Holding all other variables at their means, predicted  $CD_5$  of papers and patents increase by 303.5% and 1.3%, respectively, when the diversity of work cited increases by one standard deviation. The coefficients of the ratio of self-citations to total work cited is negative and significant for papers ( $-0.011$ ,  $p < 0.01$ ) and patents ( $-0.060$ ,  $p < 0.01$ ), showing that when researchers or inventors rely more on their own work, discovery and invention tends to be less disruptive. Again holding all other variables at their means, the predicted  $CD_5$  of papers and patents decrease by 622.9% and 18.5%, respectively, with a one standard deviation increase in the ratio. The coefficients of the interaction between mean age of work cited and dispersion in age of work cited is positive and significant for papers ( $0.000$ ,  $p < 0.01$ ) and patents ( $0.001$ ,  $p < 0.01$ ), suggesting that—holding the dispersion of the age of work cited constant—papers and patents that engage with older work are more likely to be disruptive. The predicted  $CD_5$  of papers and patents increase by a striking 2,072.4% and 58.4%, respectively, when the mean age of work cited increases by one standard deviation (about 9 and 8 years for papers and patents, respectively), again holding all other variables at their means. In summary, the regression results suggest that changes in the use of prior knowledge may contribute to the production of less disruptive science and technology.

## Data availability

Data associated with this study are freely available in a public repository at <https://doi.org/10.5281/zenodo.7258379>. Our study draws on data from six sources: the American Physical Society, JSTOR, Microsoft Academic Graph, Patents View, PubMed, and Web of Science. Data from Microsoft Academic Graph, Patents View, and PubMed are publicly available, and our repository includes complete data for analyses from these sources. Data from the American Physical Society, JSTOR, and Web of Science are not publicly available, and were used under license from their respective publishers. To facilitate replication, our repository includes limited versions of the data from these sources, which will enable calculation of basic descriptive statistics. The authors will make full versions of these data available upon request and with permission from their respective publishers.

## Code availability

Open-source code related to this study is available at <https://doi.org/10.5281/zenodo.7258379> and <http://www.cdindex.info>. We used Python v3.10.6 (pandas v1.4.3, numpy v1.23.1, matplotlib v3.5.2, seaborn v0.11.2, spacy v2.2, jupyterlab v3.4.4) to wrangle, analyze, and visualize data and to conduct statistical analyses. We used MariaDB v.10.6.4 to wrangle data. We used R v4.2.1 (ggplot2 v3.36, ggrepel v0.9.0) to visualize data. We used StataMP v17.0 (reghdfe v.5.7.3) to conduct statistical analyses.

## References

56. Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).
57. Jones, B. F., Wuchty, S. & Uzzi, B. Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).
58. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
59. Grömping, U. Estimators of relative importance in linear regression based on variance decomposition. *American Statistician* **61**, 139–147 (2007).
60. Mukherjee, S., Uzzi, B., Jones, B. & Stringer, M. A new method for identifying recombinations of existing knowledge associated with high-impact innovation. *Journal of Product Innovation Management* **33**, 224–236 (2016).
61. Christianson, N. H., Sizemore Blevins, A. & Bassett, D. S. Architecture and evolution of semantic networks in mathematics texts. *Proceedings of the Royal Society A* **476**, 20190741 (2020).
62. Newman, M. E. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences* **98**, 404–409 (2001).
63. Uzzi, B. & Spiro, J. Collaboration and creativity: The small world problem. *American Journal of Sociology* **111**, 447–504 (2005).
64. Newman, M. E. Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E* **64**, 016131 (2001).
65. Funk, R. J. Making the most of where you are: Geography, networks, and innovation in organizations. *Academy of Management Journal* **57**, 193–222 (2014).
66. Barabási, A.-L. *Network Science* 2016.
67. Blau, D. M. & Weinberg, B. A. Why the US science and engineering workforce is aging rapidly. *Proceedings of the National Academy of Sciences* **114**, 3879–3884 (2017).
68. Cui, H., Wu, L. & Evans, J. A. Aging Scientists and Slowed Advance. *arXiv preprint arXiv:2202.04044* (2022).
69. Azoulay, P., Fons-Rosen, C. & Graff Zivin, J. S. Does science advance one funeral at a time? *American Economic Review* **109**, 2889–2920 (2019).

## Acknowledgments

We thank the National Science Foundation for financial support of work related to this project (grants 1829168, 1932596, and 1829302).

## Author Contributions

R.J.F. and E.L. collaboratively contributed to the conception and design of the study. R.J.F. and M.P. collaboratively contributed to the acquisition, analysis, and interpretation of the data. R.J.F. created software used in the study. R.J.F. E.L., and M.P. collaboratively drafted and revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional Information

**Supplementary Information** is available for this paper.

**Correspondence and request for materials** should be addressed to R.J.F.

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

## Extended data table information

### Extended Data Table 1 Title:

Regression models of disruptiveness and the use of prior knowledge

### Extended Data Table 1 Footnotes:

*Notes:* This table evaluates the relationship between different measures of the use of prior scientific and technological knowledge and  $CD_5$ . Estimates are from ordinary-least-squares regressions. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Robust standard errors are shown in parentheses.

\* $p < 0.1$  \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

Disruptive  
 $CD_t = 1$

$CD_t = 0.70$   
(Axel, Wigler, & Silverstein 1984): Patent on eukaryotic cotransformation, which proposed inserting foreign genes into cells that then produce associated proteins; foundational method in biotechnology.

$CD_t = 0.62$   
(Watson & Crick 1953): Nobel-prize-winning paper that proposed a new structure of DNA; debunked the existing model of DNA structure, Pauling's 3D double helix.

$CD_t = 0.16$   
(Page 1997): Patent for Google's core algorithm to rank the importance of web pages. Presented a novel method that utilizes network theory to rank web pages based on links from other sites.

$CD_t = -0.22$   
(Kohn & Sham 1965): Nobel-Prize-winning paper that developed the Kohn-Sham equation. The equation was an application of the existing Hohenberg-Kohn theorem.

$CD_t = -0.55$   
(Baltimore 1970): Nobel-Prize-winning paper that showed that viruses with RNA genomes can be inserted into host cells; developed from existing findings on the transmission of genetic information between DNA and RNA.

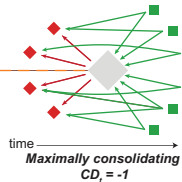
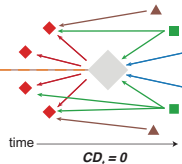
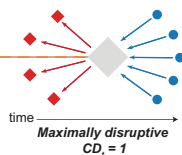
$CD_t = -0.85$   
(Hawbaker 2005): Monsanto's patent on a genetically modified soybean that is resistant to glyphosate; integrates other desirable plant characteristics such as higher yield, immunity to many diseases, and resistance to shattering.

Web of Science

USPTO

Papers

Patents



Equation

$$CD_t = \frac{1}{n} \sum_{i=1}^n -2f_{it}b_{it} + f_{it}$$

$f_{it} = 1$  if  $i$  cites the focal paper/patent; 0 if not

$b_{it} = 1$  if  $i$  cites the predecessors of the focal paper/patent; 0 if not

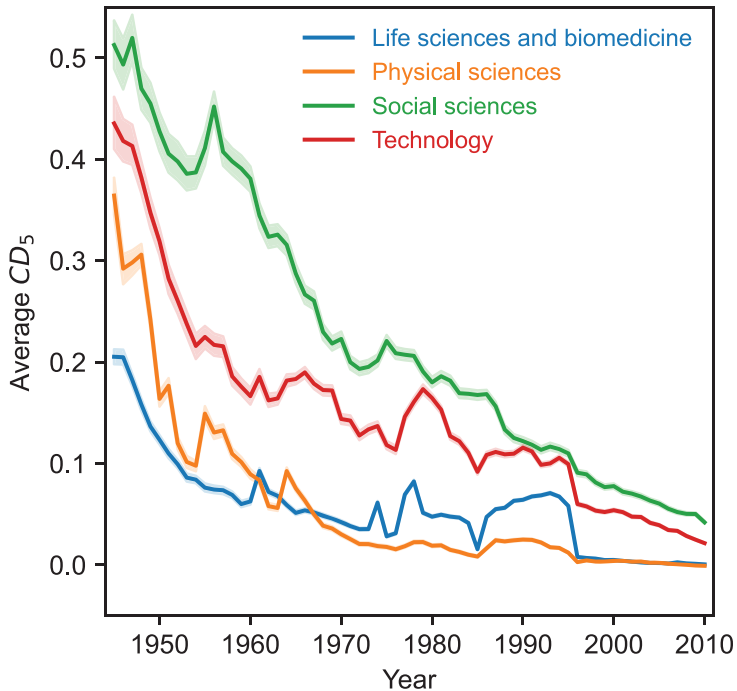
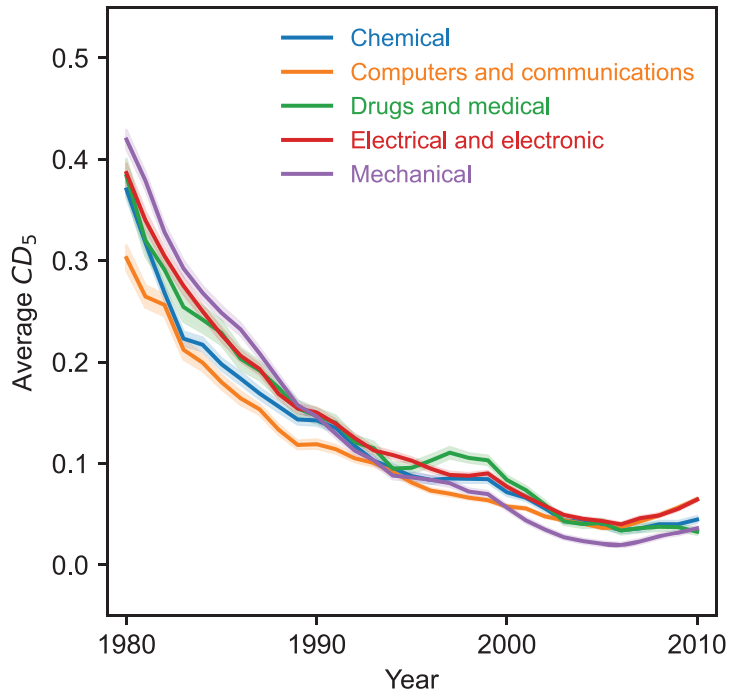
$n_t$ : number of forward cites to the focal work and/or its predecessors as of time  $t$

$$\left. \begin{aligned} (0) + (-1) + (1) + (1) + (-1) + (-1) + (0) &= 0 \\ -2(0)(1) + (0) &= 0 \\ -2(1)(1) + (1) &= -1 \\ -2(1)(0) + (1) &= 1 \\ -2(1)(0) + (1) &= 1 \\ -2(1)(1) + (1) &= -1 \\ -2(0)(1) + (0) &= 0 \end{aligned} \right\} CD_t = 0$$

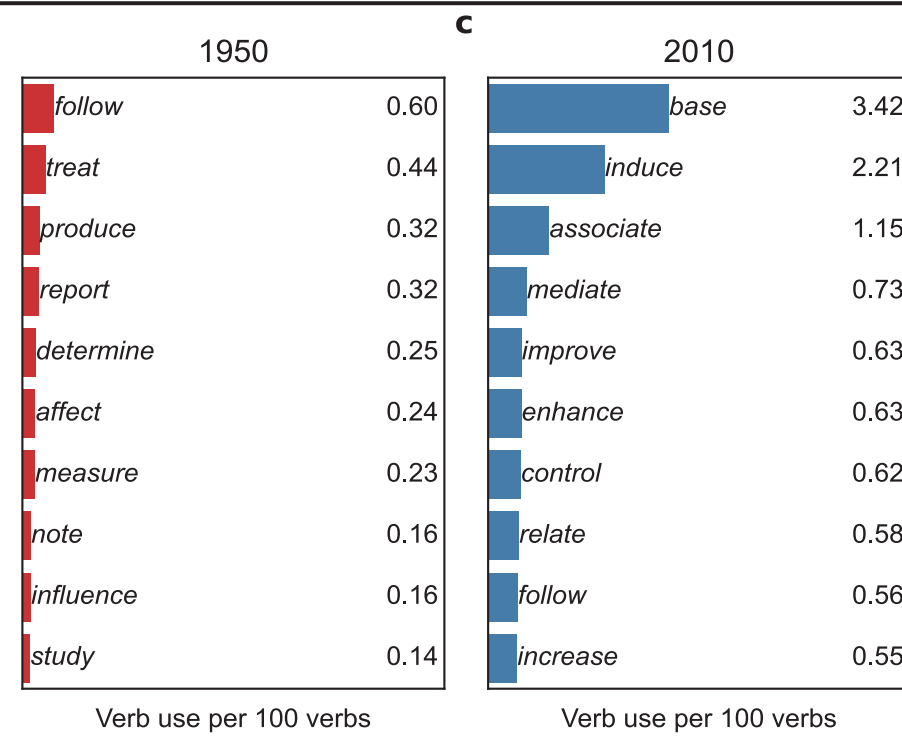
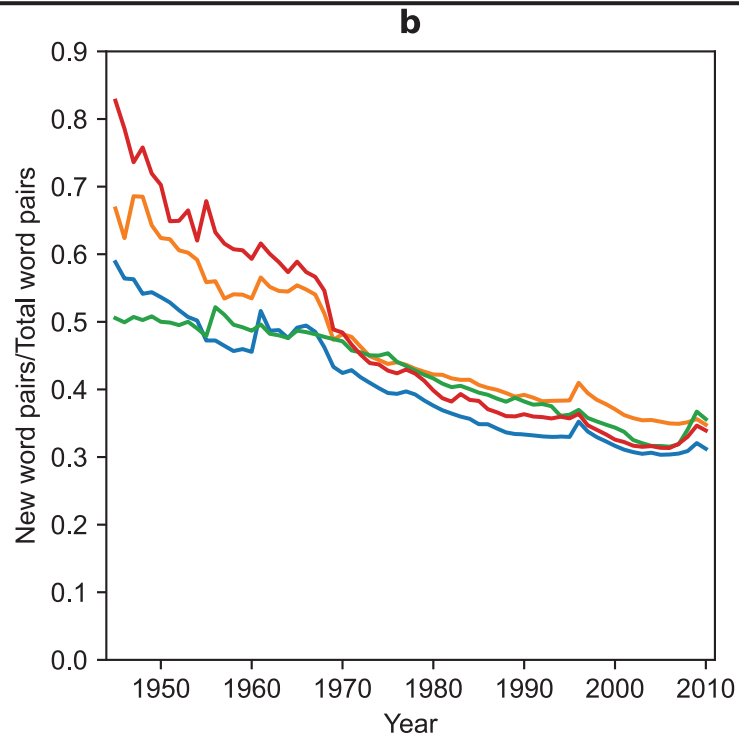
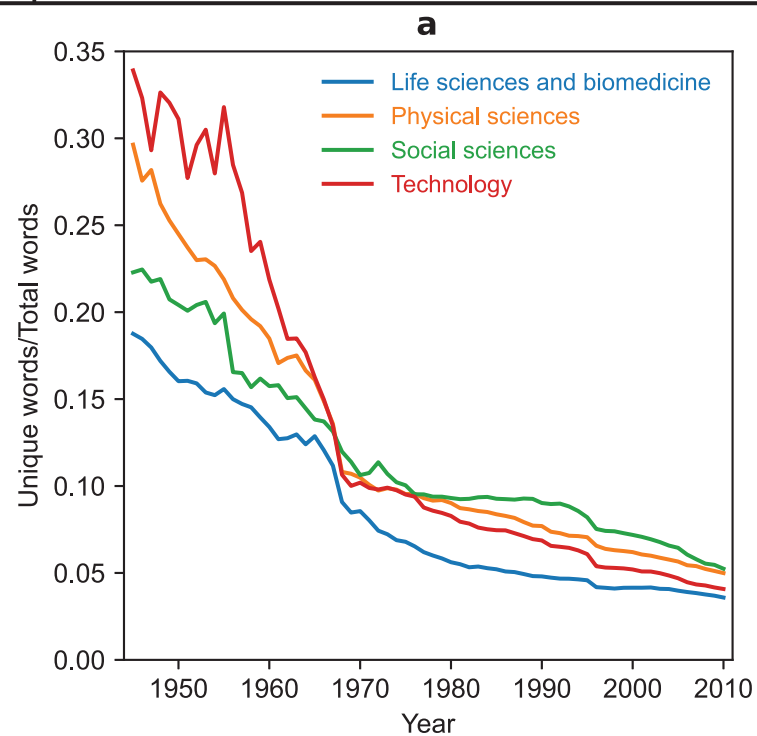
Key

- ◊ Focal paper/patent
- ◈ Focal paper/patent's predecessor
- Cites focal paper/patent (type  $f = 1$ )
- ▲ Cites predecessor (type  $b = 1$ )
- Cites focal paper/patent and predecessor (type  $f = 1, b = 1$ )

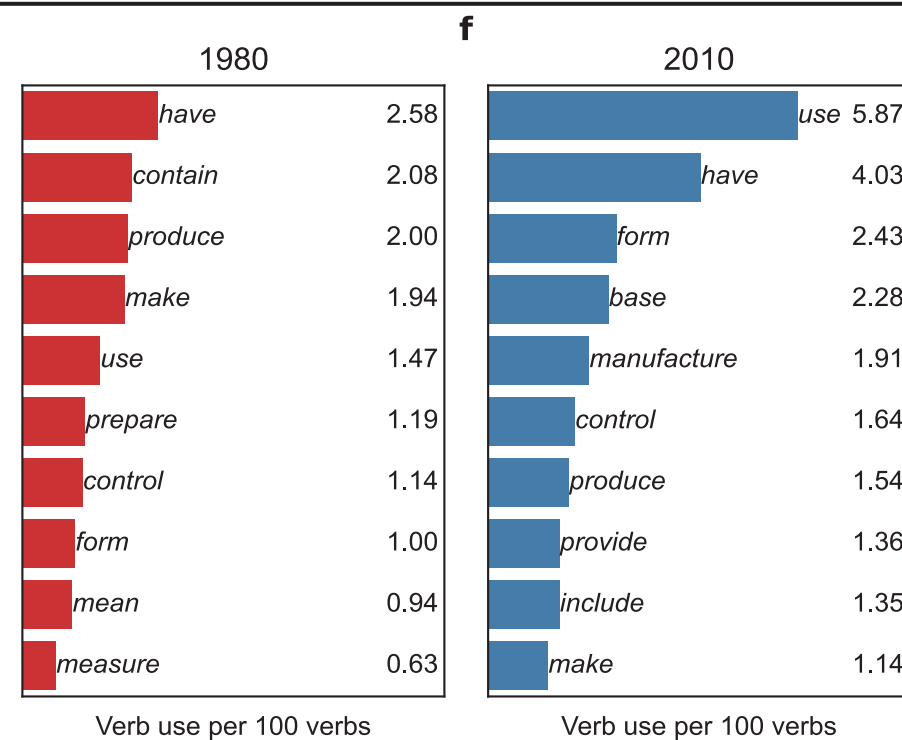
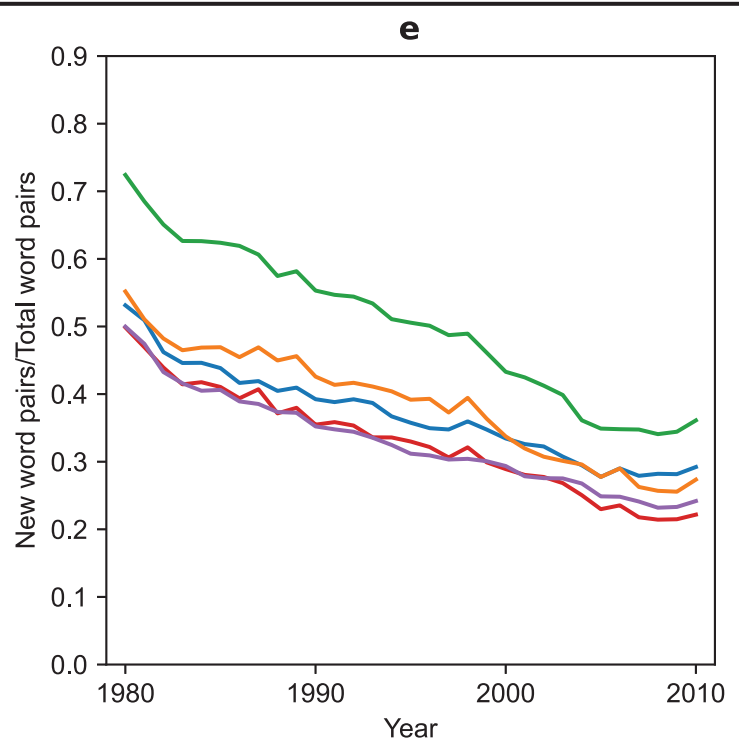
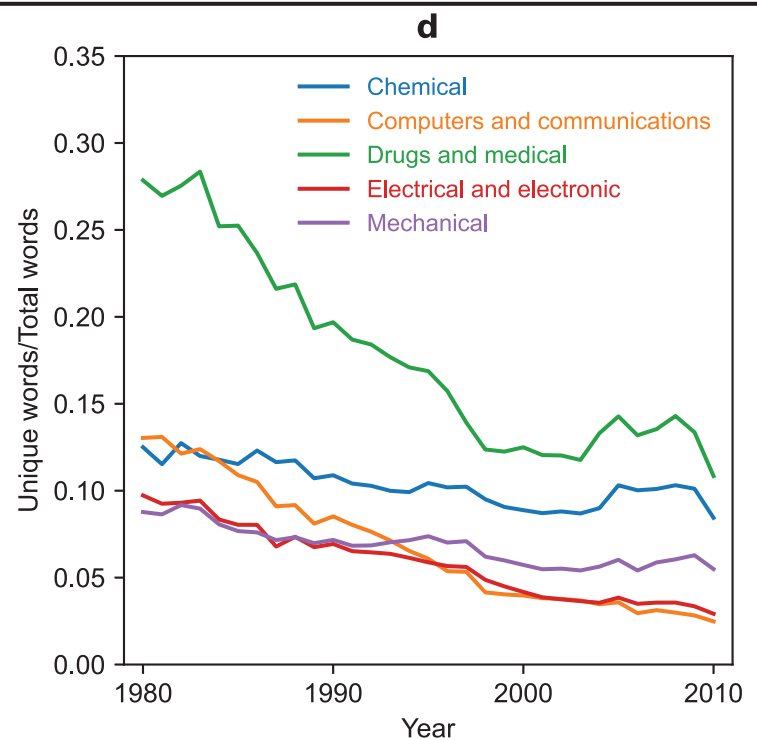
$CD_t = -1$   
Consolidating

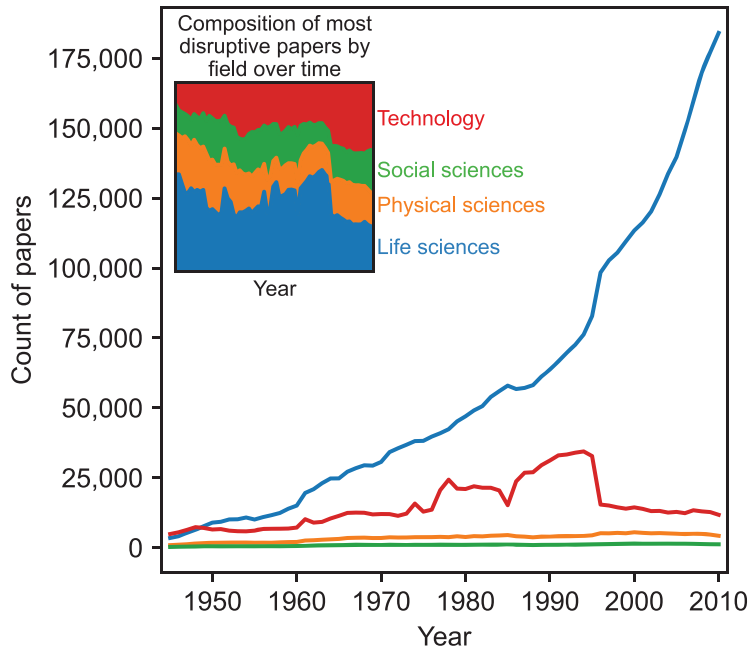
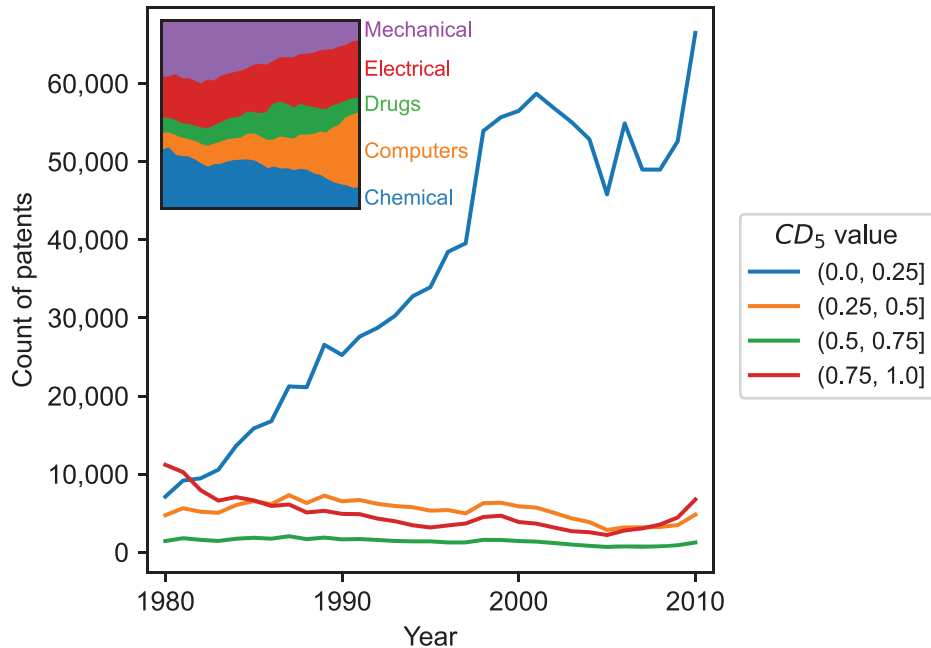
**a** Papers**b** Patents

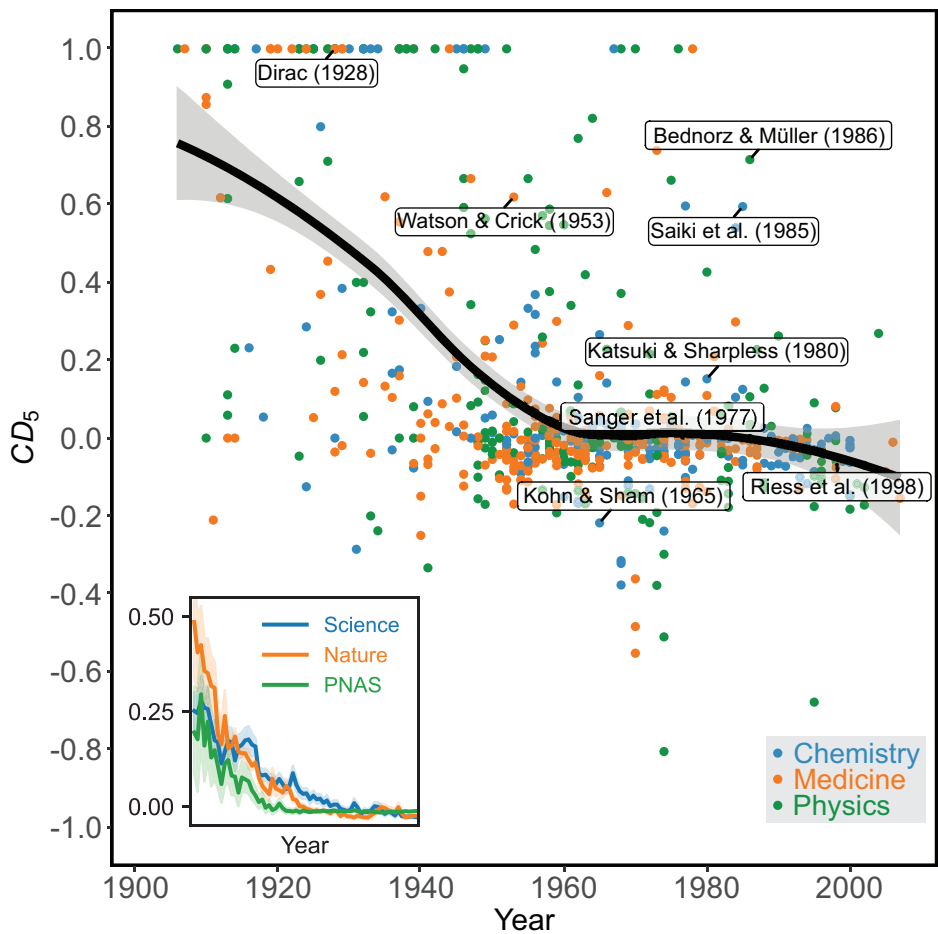
# Papers



# Patents



**a Papers****b Patents**



Dirac (1928): Discovery of the Dirac equation of relativistic quantum theory

Watson & Crick (1953): Discovery of the structure of the DNA

Kohn & Sham (1965): Development of a new method for calculating electronic structure

Sanger et al. (1977): Development of a new method for mapping the order of nucleotides

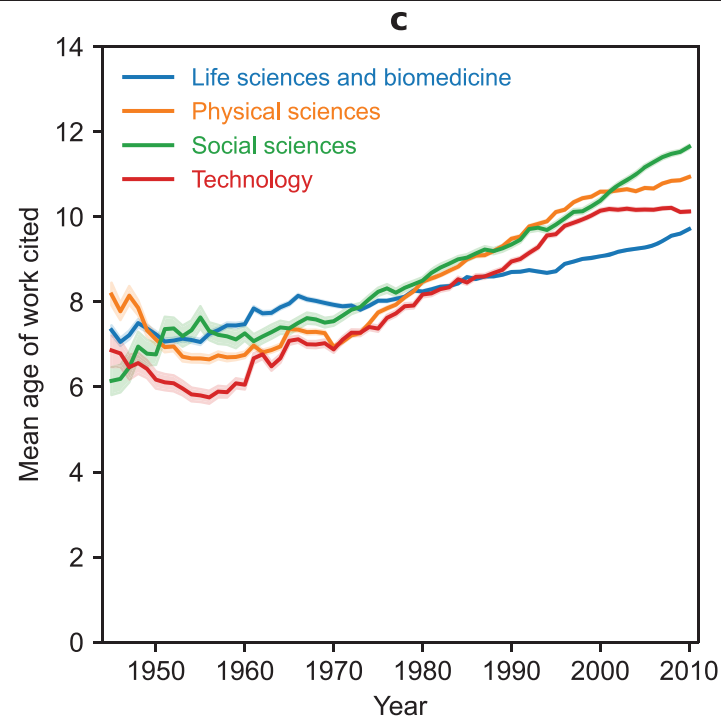
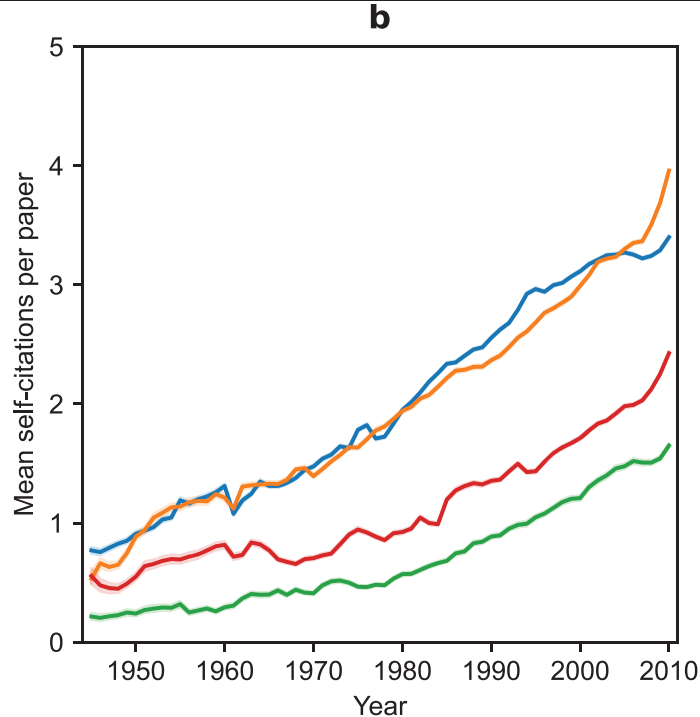
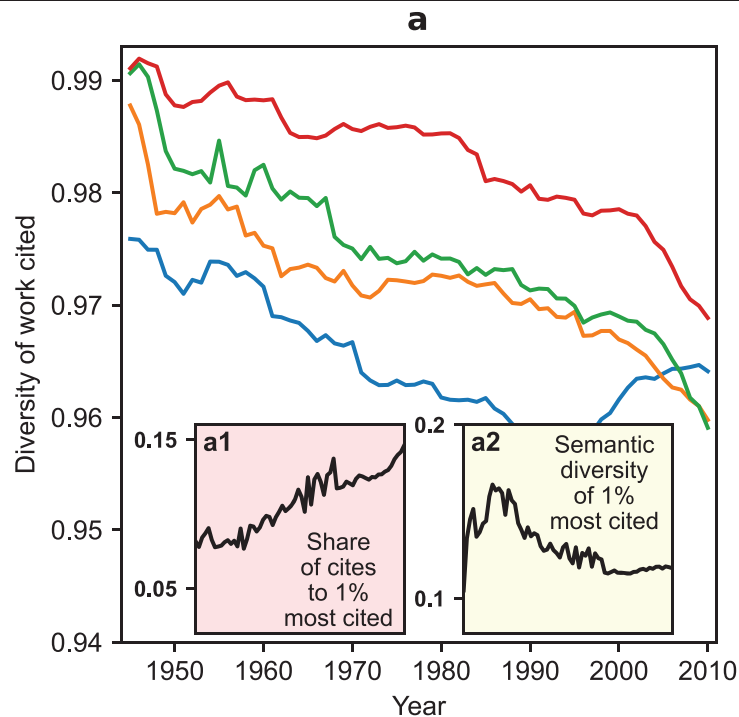
Katsuki & Sharpless (1980): Development of asymmetric epoxidation

Saiki et al. (1985): Discovery of polymerase chain reaction

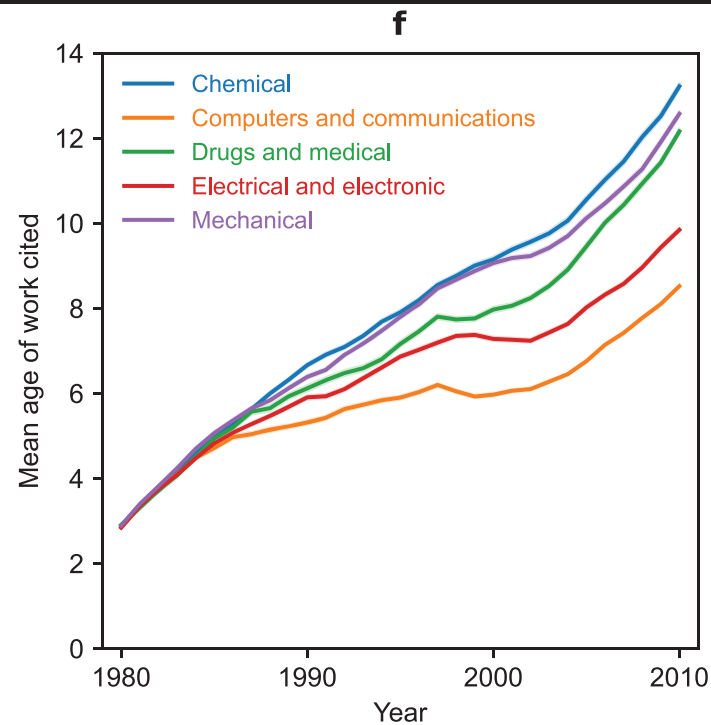
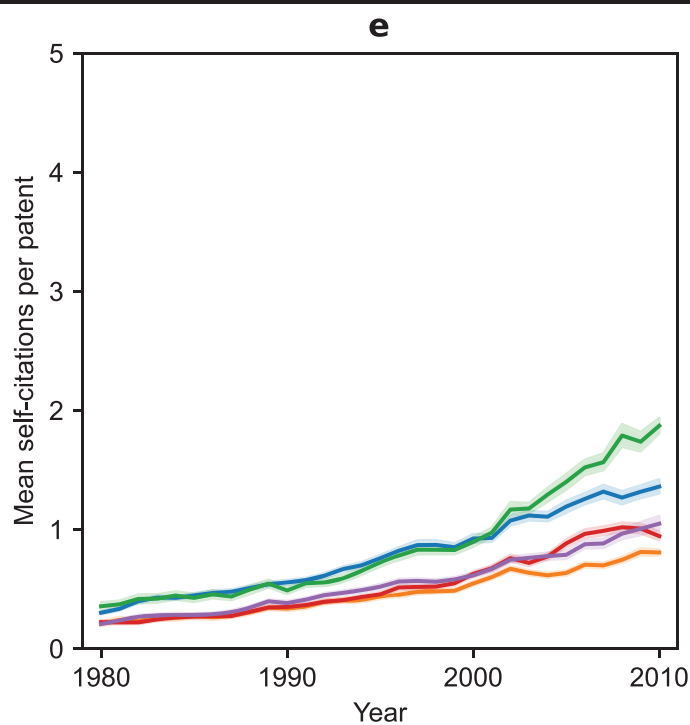
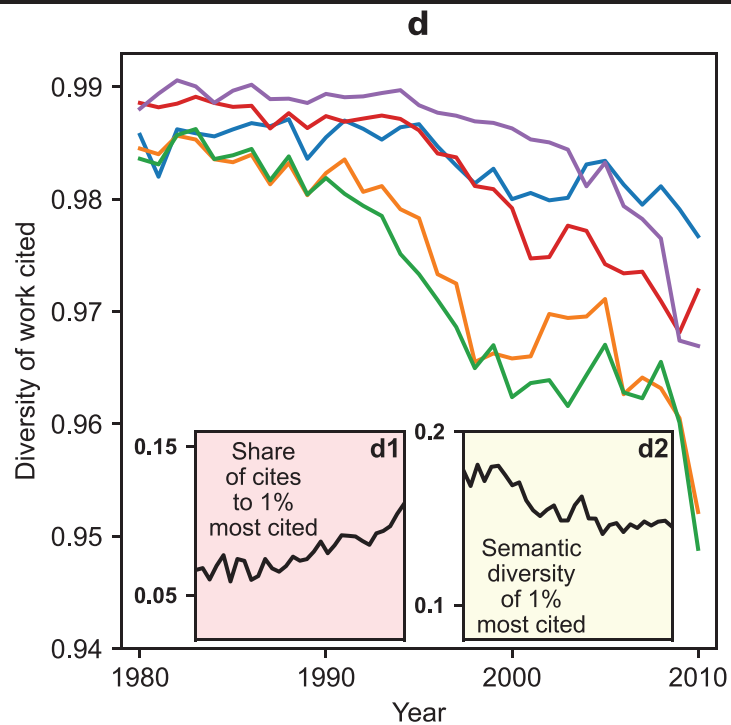
Bednorz & Müller (1986): Discovery of superconductivity in ceramic materials

Riess et al. (1998): Discovery of the accelerating expansion of the universe

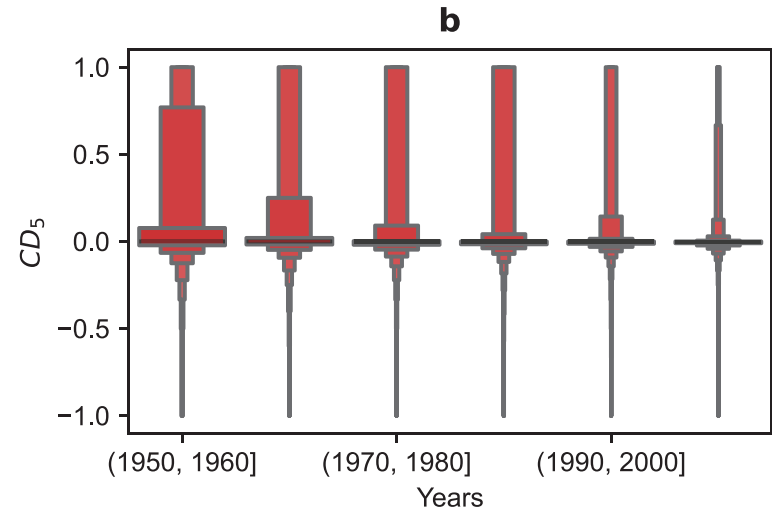
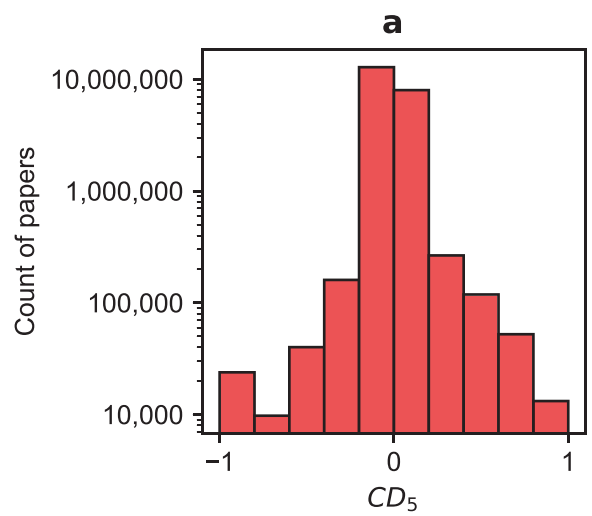
# Papers



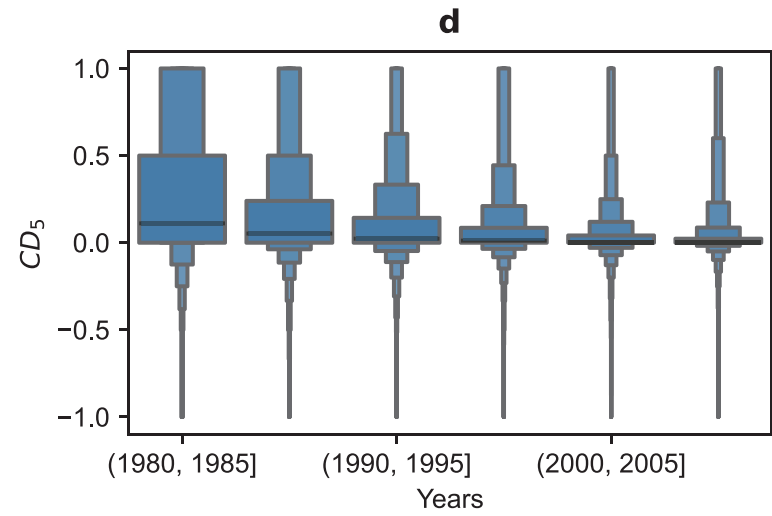
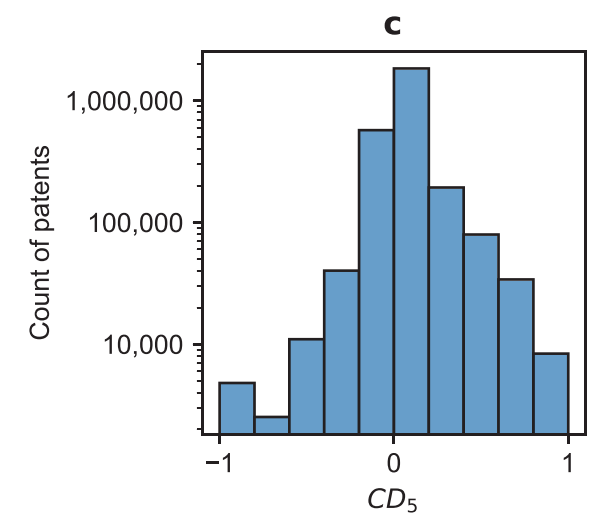
# Patents



# Papers

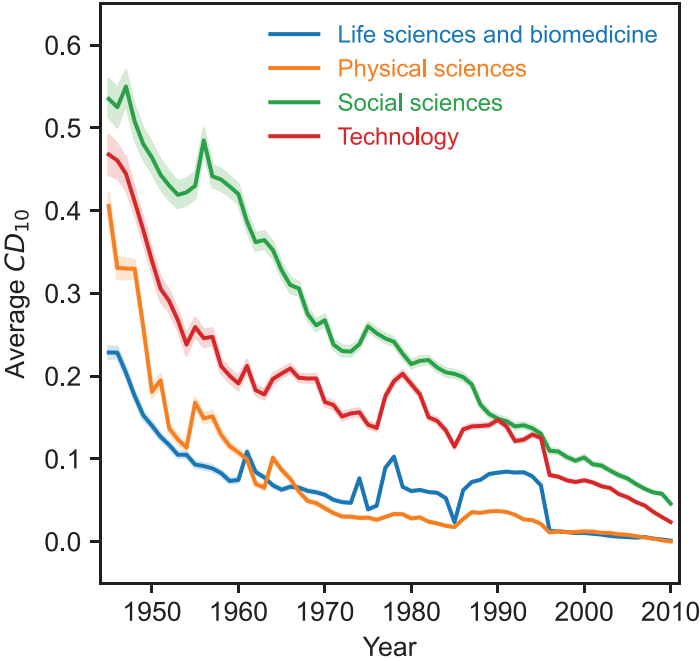


# Patents

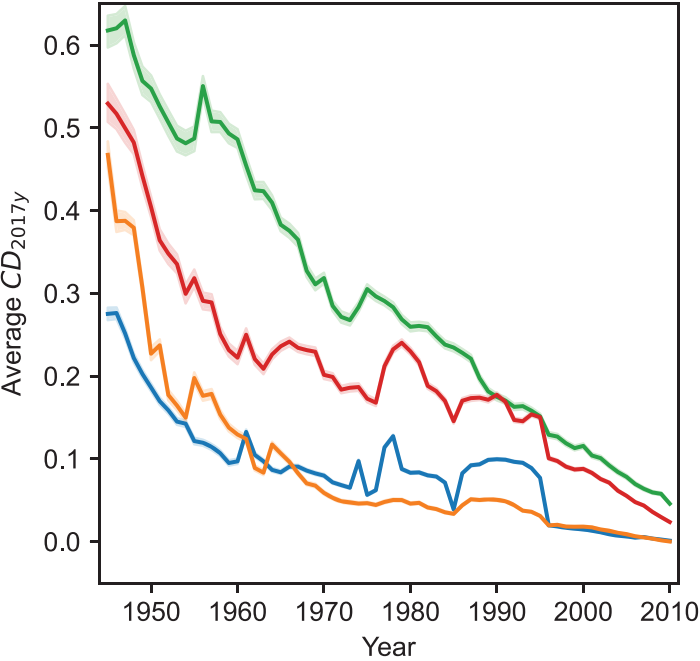


Papers

**a** CD index as of 10 years post publication

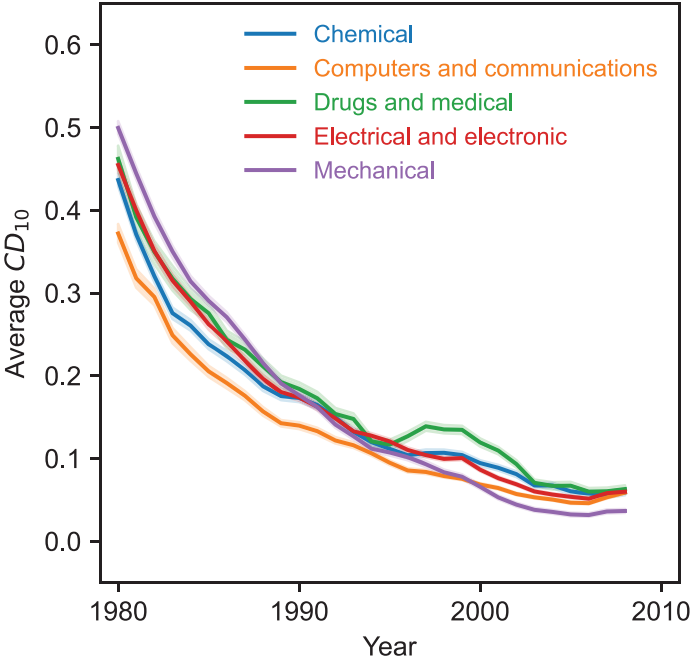


**b** CD index as of 2017

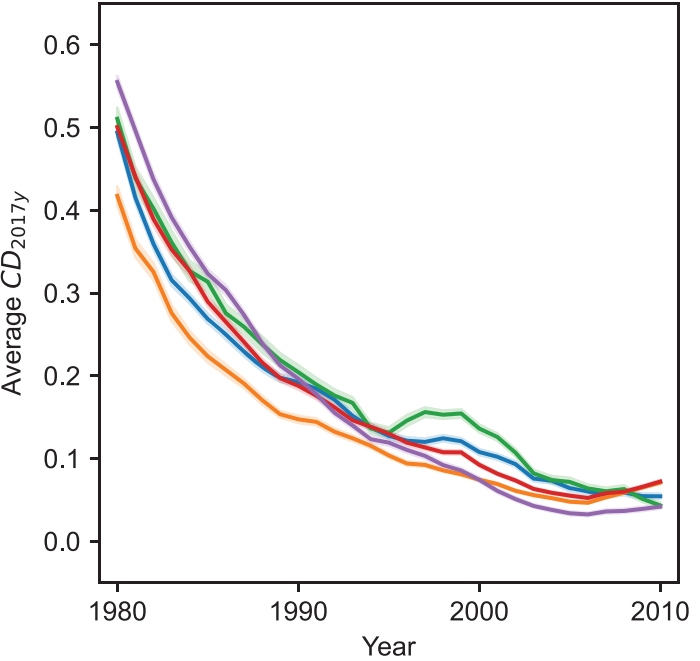


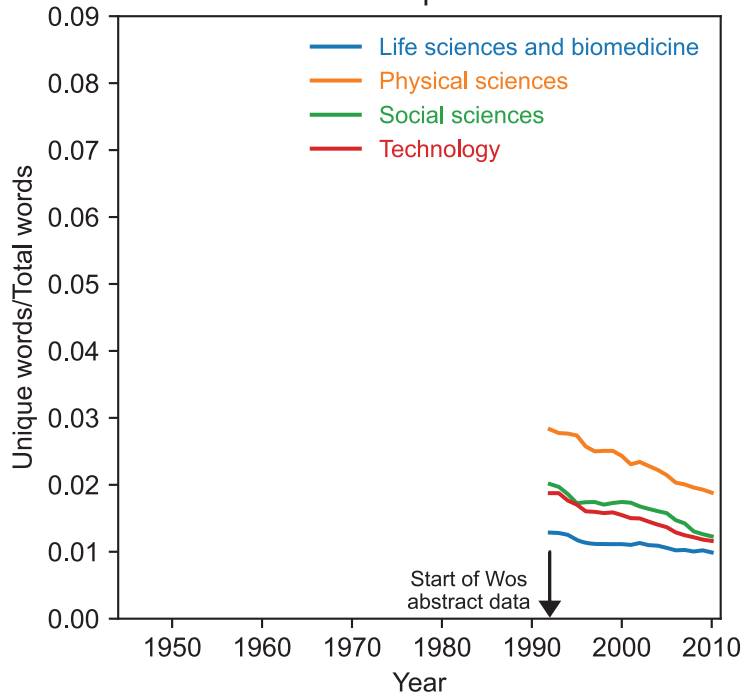
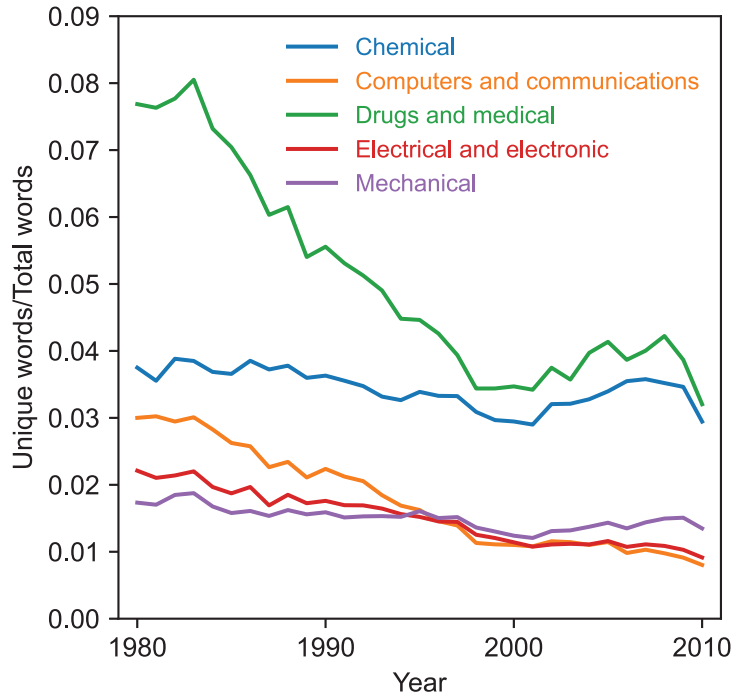
Patents

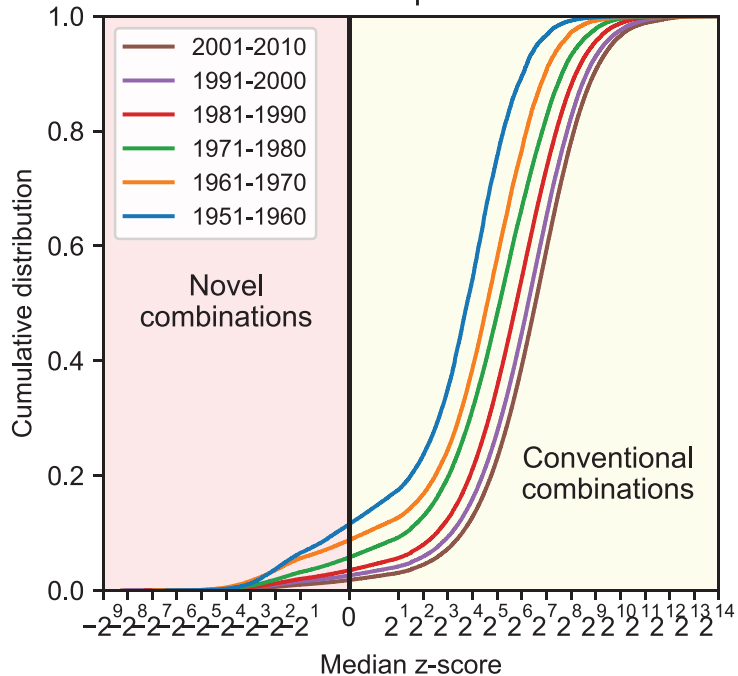
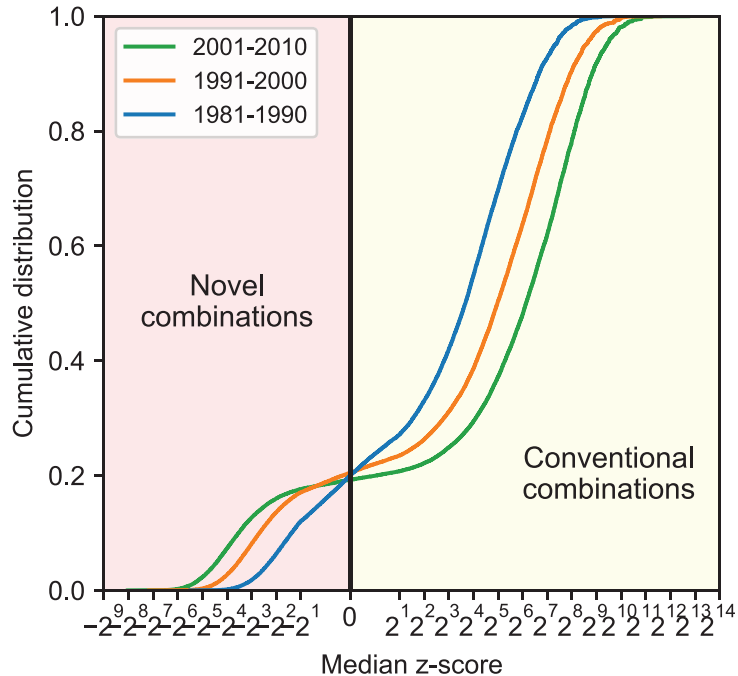
**c** CD index as of 10 years post publication

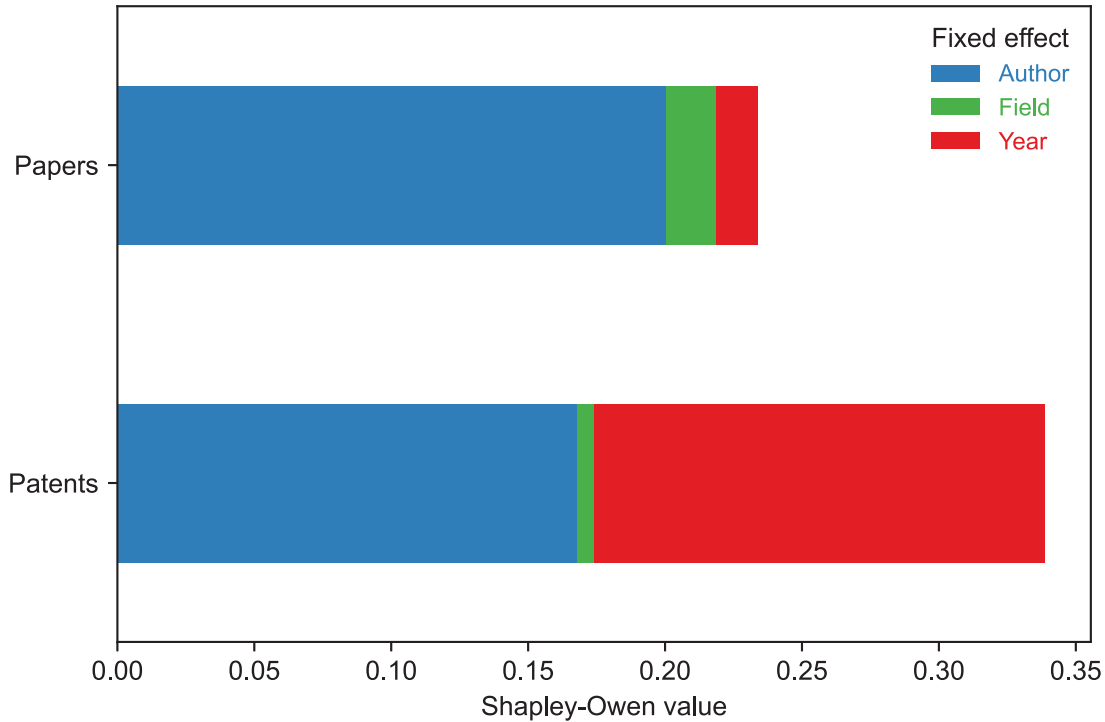


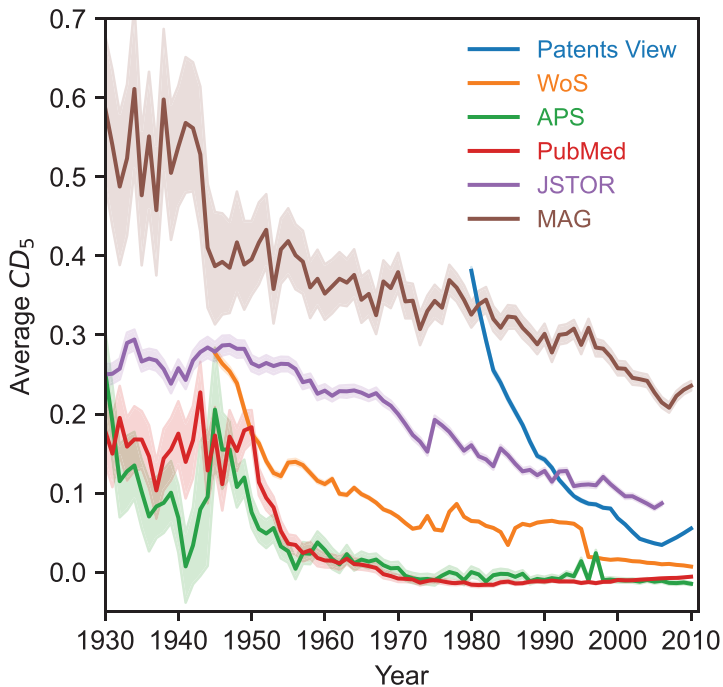
**d** CD index as of 2017

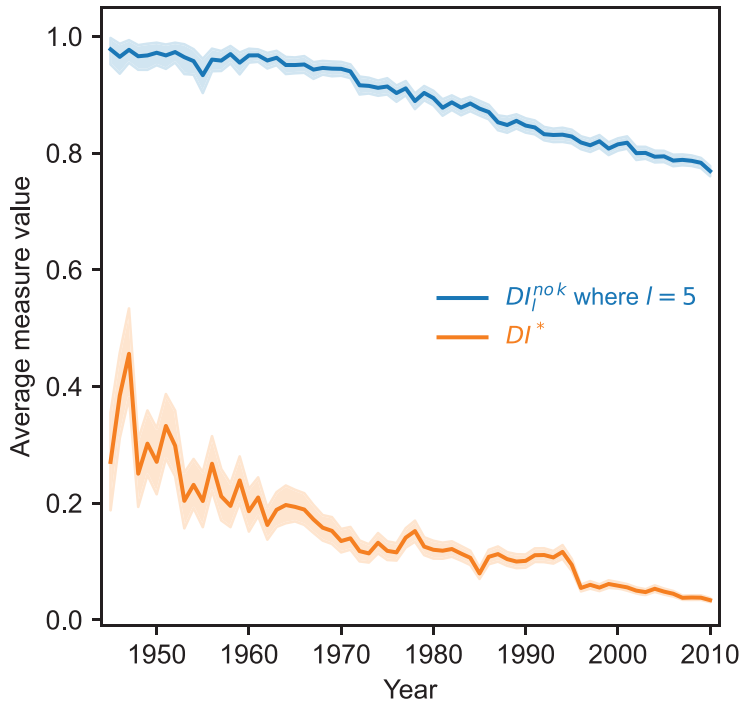
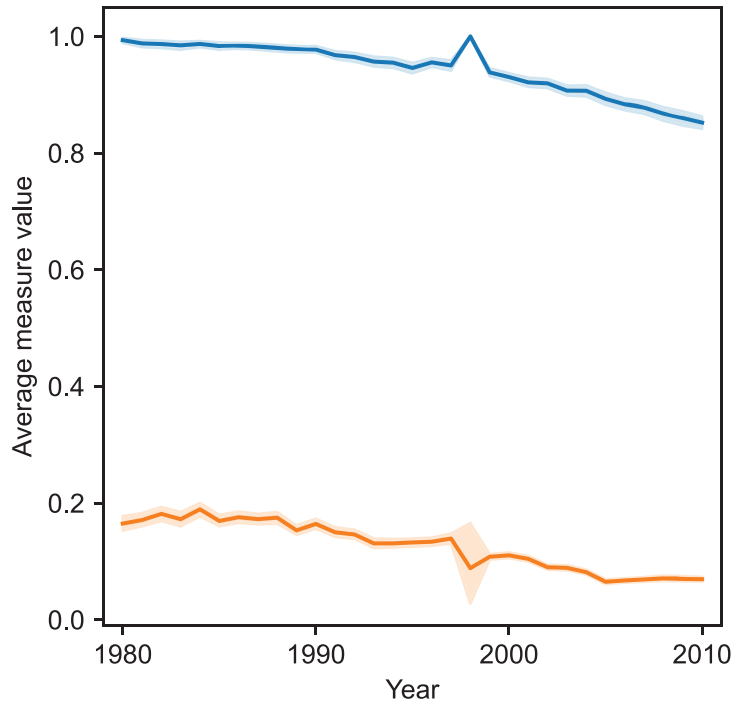


**a** Papers**b** Patents

**a** Papers**b** Patents

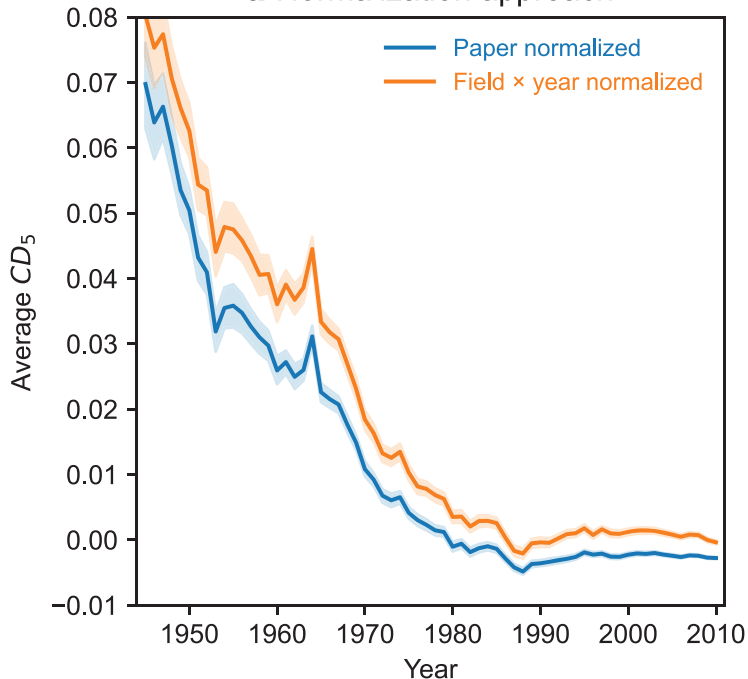




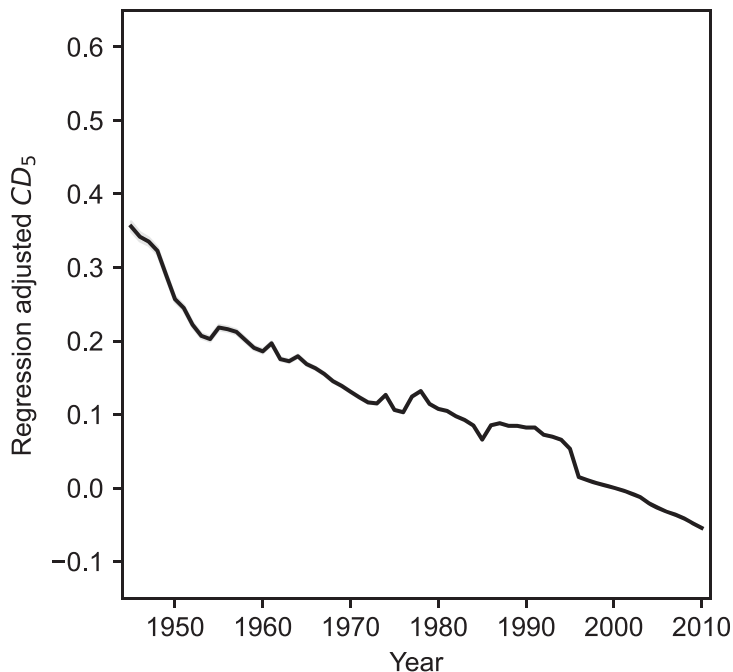
**a** Papers**b** Patents

# Papers

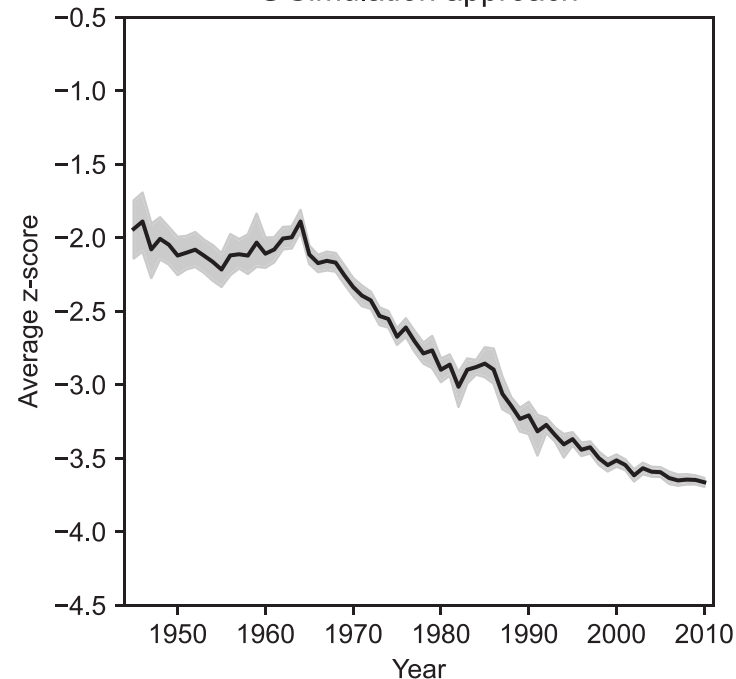
**a** Normalization approach



**b** Regression adjustment approach

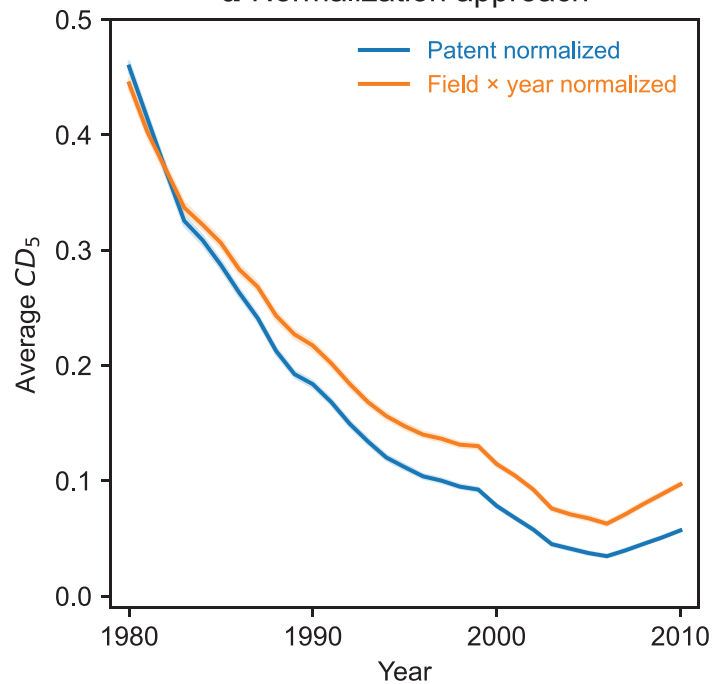


**c** Simulation approach

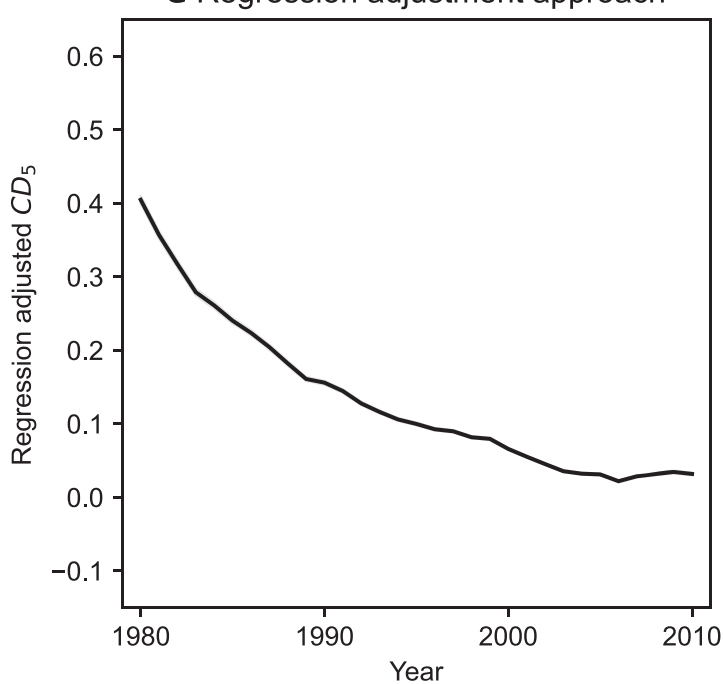


# Patents

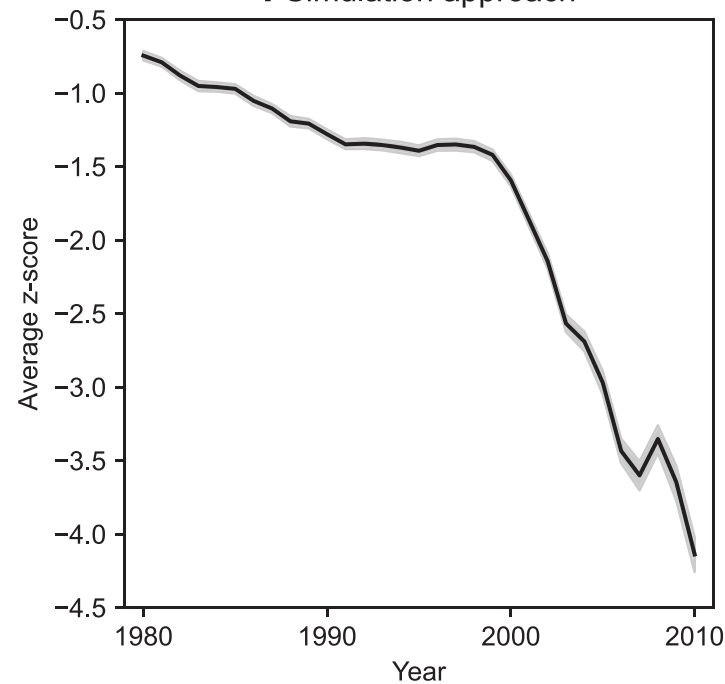
**d** Normalization approach

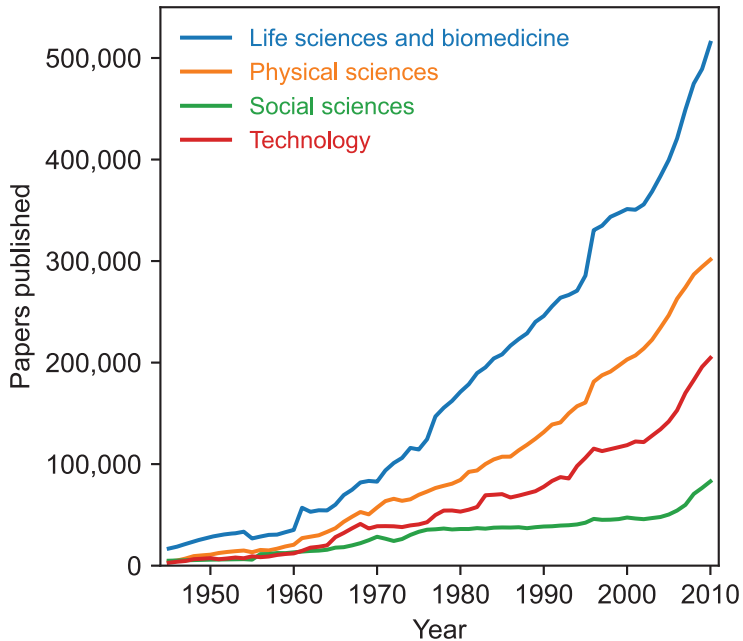
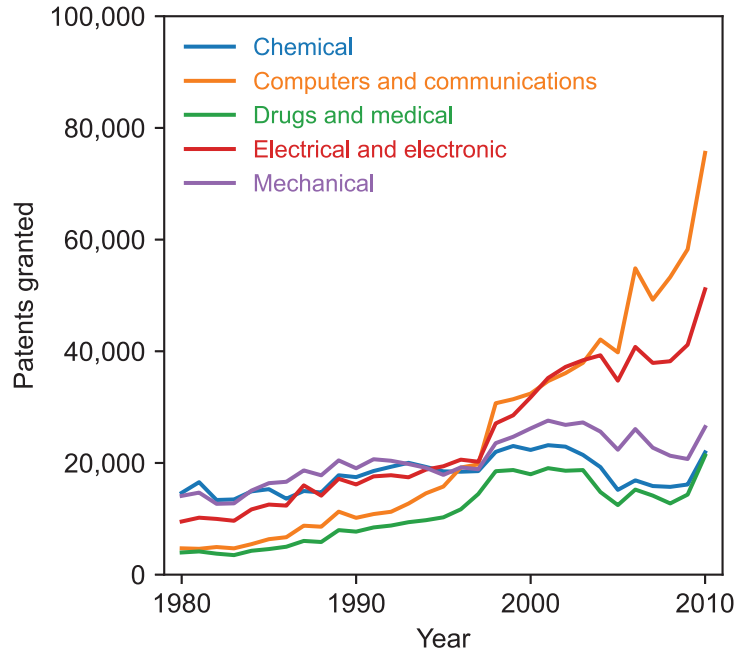


**e** Regression adjustment approach



**f** Simulation approach



**a Papers****b Patents**

|                                                          | Sample: Web of Science |                        |                        |                        |                        |                        | Sample: Patents View  |                        |                        |                        |                        |                        |
|----------------------------------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                                                          | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    | (6)                    | (7)                   | (8)                    | (9)                    | (10)                   | (11)                   | (12)                   |
| Diversity of work cited                                  | 0.3293***<br>(0.0062)  | 0.3339***<br>(0.0062)  | 0.1574***<br>(0.0061)  | 0.4578***<br>(0.0025)  | 0.1583***<br>(0.0061)  | 0.1587***<br>(0.0061)  | 0.1151***<br>(0.0158) | 0.1119***<br>(0.0158)  | 0.0873***<br>(0.0156)  | 1.3737***<br>(0.0080)  | 0.0688***<br>(0.0156)  | 0.0692***<br>(0.0156)  |
| Ratio of self-citations to total work cited              |                        | -0.0191***<br>(0.0003) | -0.0091***<br>(0.0003) | -0.0118***<br>(0.0003) | -0.0104***<br>(0.0003) | -0.0107***<br>(0.0003) |                       | -0.0606***<br>(0.0009) | -0.0557***<br>(0.0009) | -0.0671***<br>(0.0009) | -0.0585***<br>(0.0009) | -0.0597***<br>(0.0009) |
| Mean age of work cited                                   |                        |                        | 0.0034***<br>(0.0000)  | 0.0027***<br>(0.0000)  | 0.0028***<br>(0.0000)  | 0.0028***<br>(0.0000)  |                       |                        | 0.0074***<br>(0.0000)  | 0.0008***<br>(0.0001)  | 0.0046***<br>(0.0001)  | 0.0046***<br>(0.0001)  |
| Dispersion in age of work cited                          |                        |                        | -0.0051***<br>(0.0000) | -0.0069***<br>(0.0000) | -0.0063***<br>(0.0000) | -0.0063***<br>(0.0000) |                       |                        | -0.0205***<br>(0.0001) | -0.0370***<br>(0.0001) | -0.0293***<br>(0.0001) | -0.0293***<br>(0.0001) |
| Mean age of work cited × Dispersion in age of work cited |                        |                        |                        | 0.0001***<br>(0.0000)  | 0.0001***<br>(0.0000)  | 0.0001***<br>(0.0000)  |                       |                        |                        | 0.0013***<br>(0.0000)  | 0.0009***<br>(0.0000)  | 0.0009***<br>(0.0000)  |
| Mean age of team members                                 |                        |                        |                        |                        |                        | 0.0000***<br>(0.0000)  |                       |                        |                        |                        |                        | 0.0002***<br>(0.0000)  |
| Mean number of prior works produced by team members      |                        |                        |                        |                        |                        | 0.0000***<br>(0.0000)  |                       |                        |                        |                        |                        | -0.0000***<br>(0.0000) |
| Year fixed effects                                       | Yes                    | Yes                    | Yes                    | No                     | Yes                    | Yes                    | Yes                   | Yes                    | Yes                    | No                     | Yes                    | Yes                    |
| Field fixed effects                                      | Yes                    | Yes                    | Yes                    | No                     | Yes                    | Yes                    | Yes                   | Yes                    | Yes                    | No                     | Yes                    | Yes                    |
| N                                                        | 21553305               | 21553305               | 21553305               | 21553305               | 21553305               | 21553305               | 3433452               | 3433452                | 3433452                | 3433452                | 3433452                | 3433452                |
| R2                                                       | 0.02                   | 0.02                   | 0.04                   | 0.03                   | 0.04                   | 0.04                   | 0.06                  | 0.06                   | 0.10                   | 0.08                   | 0.10                   | 0.10                   |

# Contents of the supplementary information file for “Papers and patents are becoming less disruptive over time”

Michael Park, Erin Leahey, and Russell J. Funk

## **Supplementary Information Section 1**

Changes in the diversity and novelty of word use over time

## **Supplementary Information Section 2**

Changes in word use over time

## **Supplementary Information Section 3**

Disruptiveness and the growth of knowledge

## **Supplementary Table 1**

Regression models of trends in disruptiveness adjusted for publication, citation, and authorship practices

## **Supplementary Table 2**

Regression models of disruptiveness and the growth of knowledge

## **Supplementary Table 3**

Summary statistics for variables used in regression analyses

## **Supplementary references**

Supplementary information for  
“Papers and patents are becoming less disruptive over time”

Michael Park, Erin Leahey, and Russell J. Funk

## Supplementary information

### 1. Changes in the diversity and novelty of word use over time

We examined changes in the diversity and novelty of word use in papers and patents over time. Our rationale for these analyses is that decreases in disruption should be associated with decreases in the diversity of words used in science and technology. Disruptive discoveries and inventions create departures from the status quo, rendering their predecessors less useful. While this pattern alone may have the effect of reducing the diversity of words used, disruptive discoveries and inventions are also likely to introduce new words or combinations of words; part of the way that disruptive discoveries and inventions render their predecessors less useful is by introducing ideas that are more useful than those that came before, which are likely to require new words or combinations of existing words to describe. Taken together with the long memory of language (i.e., even obsolete words are still occasionally used<sup>1</sup>), we therefore anticipate a positive association between disruption and the diversity and combinatorial novelty of word use by scientists and inventors. Thus, to the extent that our observations of decreasing disruption hold, we should see a decline in the diversity of words and novelty of their pairings over time.

To evaluate for such changes, we pulled all titles and abstracts for papers and patents in our sample from Web of Science and Patents View. For titles, there was very little missing data in either Web of Science or Patents View, with titles absent in fewer than 0.01% of cases in both the former and the latter. For abstracts, Patents View also provides highly complete coverage, with only 0.32% of cases missing. Web of Science has less robust coverage of abstracts before the early 1990s; from 1945-1991, only 4.45% of papers in our sample include abstracts. Coverage is much better in later years; from 1992-2010, abstracts are included for 90.85% of papers. We therefore limit our analyses of abstract data from WoS to the 1992-2010 period.

After extracting paper and patent titles and abstracts, we completed a series of preprocess-

ing steps using spaCy, an open-source, state-of-the-art Python package for natural language processing. To begin, we tokenized each title and abstract. From the resulting lists of tokens, we then excluded those that were tagged by spaCy as stop words, tokens consisting only of digits or punctuation, and tokens that were shorter than three characters or longer than 250 characters in length. Next, we converted all remaining tokens to their lemmatized form and converted all letters to lowercase. Finally, we aggregated the resulting lists of tokens to the subfield  $\times$  year level, separately for papers and patents and for titles and abstracts.

We evaluate changes in the diversity of words used over time by computing, for each subfield  $\times$  year, the *type-token ratio*, defined as the ratio of unique words to total words. We compute this measure separately for papers and patents and for titles and abstracts, at the level of the Web of Science research area (for papers) and NBER technology category (for patents). More specifically, for each field (i.e., research area or technology category) and each year, we divide the number of unique words appearing in titles by the total number of words appearing in titles (Figure 3a and 3d). We repeat this step for the number of unique and total words appearing in abstracts (Extended Data Figure 3). The measure attains its theoretical maximum when every word is used exactly once. Thus, higher values indicate greater diversity. We find similar results when measuring unique/total words using normalized entropy.

We followed a similar approach to measure the novelty of word pairs. For each subfield  $\times$  year, we identify all pairwise combinations of words appearing in the titles of papers and patents. We then remove from that set all pairwise combinations of words in the titles of papers and patents published in the focal subfield in all previous years, which yields our count of new combinations. We then obtain our final measure by dividing this count of new combinations by the total number of word combinations made in the titles of papers and patents in the focal subfield and focal year (Figure 3b and 3e). Higher values indicate greater combinatorial novelty in word use.

## 2. Changes in word use over time

We also examined changes in the specific words used in papers and patents over time. Our rationale for these analyses is that the changes we observe in the CD index are likely to coincide with changes in approaches to discovery and invention, particularly the orientation of scientists and inventors towards prior knowledge. For example, to the extent that disruption is decreasing over time, it seems plausible that we will also observe decreases in words indicating the creation, discovery, or perception of new things. Similarly, it is also plausible that we will observe concomitant increases in the use of words that are more indicative of improvement, application, or assessment of existing things, which, consistent with the notion of consolidation, may reinforce existing streams of knowledge.

To evaluate changes in the use of specific words over time, we followed an approach similar to that described in our analyses of unique/total words, using similar samples of papers and patents and preprocessing steps. To simplify the presentation, we limit our attention to words appearing in paper and patent titles, for which, as noted previously, we have more complete data. However, the patterns we report below are also observable in analyses using paper and patent abstracts. Prior work has studied word frequencies in paper and patent titles extensively, and they are generally thought to provide a good window into the nature of science and technology<sup>2</sup>. For the present analyses, during preprocessing, we also assigned a part of speech tag to each lemma, after which we extracted all nouns, verbs, adjectives, and adverbs, which we anticipated would provide the most meaningful insights. At this stage, our data consisted of counts of lemmas by part of speech appearing in the titles of sample papers and patents. To facilitate analysis, we subsequently reshaped the data in a long-panel format, separately for papers and patents, where each row was uniquely identified by a document id  $\times$  part of speech  $\times$  token.

We then examined changes in the top 10 most frequently used words in paper and patent titles by decade. For patents, we present these word frequencies for the years 1980 and 2010;

for papers, our time series is longer, and therefore we present frequencies for 1950 and 2010. Analyses (available upon request from the authors) that include additional years (e.g., for each decade) yielded consistent results.

To simplify the presentation and conserve space, we focus our reporting on the results for verbs, which also generally yielded more noteworthy patterns (the most frequent nouns were often topical in nature; the most frequent adverbs and adjectives tended to be general/stop words).

We find evidence of a qualitative shift in word use that is consistent with our quantitative findings on the decline in disruptive activity using the CD index. Figure 3 shows the most common verbs in paper (Figure 3c) and patent titles (Figure 3f) in the first and last decade of each sample. In earlier decades, for example, verbs evoking creation (e.g., “produce”, “form”, “prepare”, “make”), discovery (e.g., “determine”, “report”), and perception (e.g., “measure”) are prevalent in both paper and patent titles. In later decades, these verbs are almost completely displaced by those more evocative of the improvement (e.g., “improve”, “enhance”, “increase”), application (e.g., “use”, “include”), or assessment (e.g., “associate”, “mediate”, “relate”) of existing scientific and technological knowledge and artifacts. Thus, we observe a decrease in the use of verbs indicative of processes of disruption and a simultaneous increase in the use of verbs indicative of processes of consolidation.

These results are especially noteworthy when recalling that they are based on raw data, with no adjustment or transformation other than basic text preprocessing. Overall, then, the patterns offer compelling support for the findings we observe on the changing nature of science and technology using the CD index.

### 3. Disruptiveness and the growth of knowledge

Extended Data Figure 9 plots the number of new papers published and archived by WoS (panel a) and utility patents granted by the United States Patent and Trademark Office each year (panel b). The figure shows that there has been a sharp and consistent increase in the number of new papers and patents. The rate of new papers and patents added year by year seems to be accelerating, expanding the existing stock of knowledge rapidly, and thereby potentially placing a “burden of knowledge” on scientists and inventors.

Models in Supplementary Table 2 evaluate the relationship between the growth of knowledge and disruptiveness. Models 1 and 4 proxy for new knowledge based on the number of new works (patents or papers) produced in the focal year. Models 2 and 5 proxy for new knowledge based on the number of new works (patents or papers) produced in the five most recent years. Models 3 and 6 proxy for new knowledge based on the number of new works (patents or papers) produced in the ten most recent years. We find divergent results for papers and patents; for the former, there is a negative association between new knowledge and  $CD_5$ ; for the latter, the association is positive. This divergent pattern motivates our subsequent analyses (Extended Data Table 1), on the use of prior knowledge and disruptiveness.

**Supplementary Table 1: Regression models of trends in disruptiveness adjusted for publication, citation, and authorship practices**

|                             | Sample: Web of Science |      |          |      |          |      |          |      | Sample: Patents View |      |          |      |          |      |          |      |
|-----------------------------|------------------------|------|----------|------|----------|------|----------|------|----------------------|------|----------|------|----------|------|----------|------|
|                             | (1)                    |      | (2)      |      | (3)      |      | (4)      |      | (5)                  |      | (6)      |      | (7)      |      | (8)      |      |
|                             | b                      | se   | b        | se   | b        | se   | b        | se   | b                    | se   | b        | se   | b        | se   | b        | se   |
| Year=1946                   | -0.01***               | 0.00 | -0.01*** | 0.00 | -0.01*** | 0.00 | -0.01*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1947                   | -0.02***               | 0.00 | -0.02*** | 0.00 | -0.02*** | 0.00 | -0.02*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1948                   | -0.04***               | 0.00 | -0.03*** | 0.00 | -0.03*** | 0.00 | -0.03*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1949                   | -0.07***               | 0.00 | -0.07*** | 0.00 | -0.07*** | 0.00 | -0.07*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1950                   | -0.10***               | 0.00 | -0.10*** | 0.00 | -0.10*** | 0.00 | -0.10*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1951                   | -0.11***               | 0.00 | -0.11*** | 0.00 | -0.11*** | 0.00 | -0.11*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1952                   | -0.14***               | 0.00 | -0.13*** | 0.00 | -0.13*** | 0.00 | -0.13*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1953                   | -0.15***               | 0.00 | -0.15*** | 0.00 | -0.14*** | 0.00 | -0.15*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1954                   | -0.16***               | 0.00 | -0.15*** | 0.00 | -0.15*** | 0.00 | -0.15*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1955                   | -0.14***               | 0.00 | -0.13*** | 0.00 | -0.13*** | 0.00 | -0.14*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1956                   | -0.14***               | 0.00 | -0.13*** | 0.00 | -0.13*** | 0.00 | -0.14*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1957                   | -0.14***               | 0.00 | -0.14*** | 0.00 | -0.13*** | 0.00 | -0.14*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1958                   | -0.15***               | 0.00 | -0.15*** | 0.00 | -0.14*** | 0.00 | -0.15*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1959                   | -0.16***               | 0.00 | -0.16*** | 0.00 | -0.15*** | 0.00 | -0.17*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1960                   | -0.17***               | 0.00 | -0.16*** | 0.00 | -0.16*** | 0.00 | -0.17*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1961                   | -0.16***               | 0.00 | -0.15*** | 0.00 | -0.15*** | 0.00 | -0.16*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1962                   | -0.18***               | 0.00 | -0.17*** | 0.00 | -0.16*** | 0.00 | -0.18*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1963                   | -0.18***               | 0.00 | -0.17*** | 0.00 | -0.17*** | 0.00 | -0.18*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1964                   | -0.17***               | 0.00 | -0.17*** | 0.00 | -0.16*** | 0.00 | -0.18*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1965                   | -0.18***               | 0.00 | -0.18*** | 0.00 | -0.17*** | 0.00 | -0.19*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1966                   | -0.18***               | 0.00 | -0.18*** | 0.00 | -0.17*** | 0.00 | -0.19*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1967                   | -0.19***               | 0.00 | -0.19*** | 0.00 | -0.18*** | 0.00 | -0.20*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1968                   | -0.20***               | 0.00 | -0.20*** | 0.00 | -0.18*** | 0.00 | -0.21*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1969                   | -0.20***               | 0.00 | -0.20*** | 0.00 | -0.19*** | 0.00 | -0.22*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1970                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.20*** | 0.00 | -0.22*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1971                   | -0.22***               | 0.00 | -0.21*** | 0.00 | -0.20*** | 0.00 | -0.23*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1972                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.21*** | 0.00 | -0.24*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1973                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.20*** | 0.00 | -0.24*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1974                   | -0.21***               | 0.00 | -0.20*** | 0.00 | -0.19*** | 0.00 | -0.23*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1975                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.21*** | 0.00 | -0.25*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1976                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.21*** | 0.00 | -0.25*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1977                   | -0.20***               | 0.00 | -0.20*** | 0.00 | -0.19*** | 0.00 | -0.23*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1978                   | -0.19***               | 0.00 | -0.19*** | 0.00 | -0.18*** | 0.00 | -0.22*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1979                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.24*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1980                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.25*** | 0.00 |                      |      |          |      |          |      |          |      |
| Year=1981                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.25*** | 0.00 | -0.05***             | 0.00 | -0.05*** | 0.00 | -0.05*** | 0.00 | -0.05*** | 0.00 |
| Year=1982                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.20*** | 0.00 | -0.26*** | 0.00 | -0.09***             | 0.00 | -0.09*** | 0.00 | -0.09*** | 0.00 | -0.09*** | 0.00 |
| Year=1983                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.20*** | 0.00 | -0.26*** | 0.00 | -0.13***             | 0.00 | -0.13*** | 0.00 | -0.12*** | 0.00 | -0.13*** | 0.00 |
| Year=1984                   | -0.23***               | 0.00 | -0.23*** | 0.00 | -0.21*** | 0.00 | -0.27*** | 0.00 | -0.14***             | 0.00 | -0.14*** | 0.00 | -0.14*** | 0.00 | -0.14*** | 0.00 |
| Year=1985                   | -0.24***               | 0.00 | -0.24*** | 0.00 | -0.22*** | 0.00 | -0.29*** | 0.00 | -0.16***             | 0.00 | -0.16*** | 0.00 | -0.16*** | 0.00 | -0.16*** | 0.00 |
| Year=1986                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.20*** | 0.00 | -0.27*** | 0.00 | -0.18***             | 0.00 | -0.18*** | 0.00 | -0.18*** | 0.00 | -0.18*** | 0.00 |
| Year=1987                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.19*** | 0.00 | -0.27*** | 0.00 | -0.19***             | 0.00 | -0.19*** | 0.00 | -0.19*** | 0.00 | -0.20*** | 0.00 |
| Year=1988                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.19*** | 0.00 | -0.27*** | 0.00 | -0.22***             | 0.00 | -0.22*** | 0.00 | -0.21*** | 0.00 | -0.22*** | 0.00 |
| Year=1989                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.27*** | 0.00 | -0.23***             | 0.00 | -0.23*** | 0.00 | -0.23*** | 0.00 | -0.24*** | 0.00 |
| Year=1990                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.27*** | 0.00 | -0.24***             | 0.00 | -0.24*** | 0.00 | -0.23*** | 0.00 | -0.25*** | 0.00 |
| Year=1991                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.18*** | 0.00 | -0.27*** | 0.00 | -0.25***             | 0.00 | -0.25*** | 0.00 | -0.25*** | 0.00 | -0.26*** | 0.00 |
| Year=1992                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.19*** | 0.00 | -0.28*** | 0.00 | -0.26***             | 0.00 | -0.26*** | 0.00 | -0.26*** | 0.00 | -0.28*** | 0.00 |
| Year=1993                   | -0.21***               | 0.00 | -0.21*** | 0.00 | -0.18*** | 0.00 | -0.29*** | 0.00 | -0.27***             | 0.00 | -0.27*** | 0.00 | -0.27*** | 0.00 | -0.29*** | 0.00 |
| Year=1994                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.19*** | 0.00 | -0.29*** | 0.00 | -0.29***             | 0.00 | -0.28*** | 0.00 | -0.28*** | 0.00 | -0.30*** | 0.00 |
| Year=1995                   | -0.22***               | 0.00 | -0.22*** | 0.00 | -0.19*** | 0.00 | -0.30*** | 0.00 | -0.29***             | 0.00 | -0.29*** | 0.00 | -0.28*** | 0.00 | -0.30*** | 0.00 |
| Year=1996                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.34*** | 0.00 | -0.29***             | 0.00 | -0.29*** | 0.00 | -0.29*** | 0.00 | -0.31*** | 0.00 |
| Year=1997                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.35*** | 0.00 | -0.30***             | 0.00 | -0.30*** | 0.00 | -0.29*** | 0.00 | -0.31*** | 0.00 |
| Year=1998                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.35*** | 0.00 | -0.30***             | 0.00 | -0.30*** | 0.00 | -0.29*** | 0.00 | -0.32*** | 0.00 |
| Year=1999                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.35*** | 0.00 | -0.30***             | 0.00 | -0.30*** | 0.00 | -0.29*** | 0.00 | -0.33*** | 0.00 |
| Year=2000                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.36*** | 0.00 | -0.31***             | 0.00 | -0.31*** | 0.00 | -0.31*** | 0.00 | -0.34*** | 0.00 |
| Year=2001                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.36*** | 0.00 | -0.32***             | 0.00 | -0.32*** | 0.00 | -0.31*** | 0.00 | -0.35*** | 0.00 |
| Year=2002                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.36*** | 0.00 | -0.33***             | 0.00 | -0.33*** | 0.00 | -0.32*** | 0.00 | -0.36*** | 0.00 |
| Year=2003                   | -0.26***               | 0.00 | -0.26*** | 0.00 | -0.23*** | 0.00 | -0.37*** | 0.00 | -0.34***             | 0.00 | -0.34*** | 0.00 | -0.33*** | 0.00 | -0.37*** | 0.00 |
| Year=2004                   | -0.27***               | 0.00 | -0.27*** | 0.00 | -0.23*** | 0.00 | -0.38*** | 0.00 | -0.34***             | 0.00 | -0.34*** | 0.00 | -0.33*** | 0.00 | -0.37*** | 0.00 |
| Year=2005                   | -0.27***               | 0.00 | -0.27*** | 0.00 | -0.23*** | 0.00 | -0.38*** | 0.00 | -0.34***             | 0.00 | -0.35*** | 0.00 | -0.33*** | 0.00 | -0.37*** | 0.00 |
| Year=2006                   | -0.27***               | 0.00 | -0.27*** | 0.00 | -0.23*** | 0.00 | -0.39*** | 0.00 | -0.35***             | 0.00 | -0.35*** | 0.00 | -0.33*** | 0.00 | -0.38*** | 0.00 |
| Year=2007                   | -0.27***               | 0.00 | -0.27*** | 0.00 | -0.23*** | 0.00 | -0.39*** | 0.00 | -0.34***             | 0.00 | -0.34*** | 0.00 | -0.33*** | 0.00 | -0.38*** | 0.00 |
| Year=2008                   | -0.27***               | 0.00 | -0.27*** | 0.00 | -0.23*** | 0.00 | -0.40*** | 0.00 | -0.34***             | 0.00 | -0.34*** | 0.00 | -0.32*** | 0.00 | -0.37*** | 0.00 |
| Year=2009                   | -0.27***               | 0.00 | -0.28*** | 0.00 | -0.23*** | 0.00 | -0.40*** | 0.00 | -0.33***             | 0.00 | -0.33*** | 0.00 | -0.32*** | 0.00 | -0.37*** | 0.00 |
| Year=2010                   | -0.27***               | 0.00 | -0.28*** | 0.00 | -0.22*** | 0.00 | -0.41*** | 0.00 | -0.33***             | 0.00 | -0.33*** | 0.00 | -0.31*** | 0.00 | -0.37*** | 0.00 |
| Constant                    | 0.28***                | 0.00 | 0.28***  | 0.00 | 0.30***  | 0.00 | 0.23***  | 0.00 | 0.38***              | 0.00 | 0.38***  | 0.00 | 0.38***  | 0.00 | 0.32***  | 0.00 |
| Subfield fixed effects      | No                     |      | Yes      |      | Yes      |      | Yes      |      | No                   |      | Yes      |      | Yes      |      | Yes      |      |
| Paper/patent-level controls | No                     |      | No       |      | Yes      |      | Yes      |      | No                   |      | No       |      | Yes      |      | Yes      |      |
| Field × year-level controls | No                     |      | No       |      | No       |      | Yes      |      | No                   |      | No       |      | No       |      | Yes      |      |
| N                           | 22479429               |      | 22456096 |      | 22456096 |      | 22456096 |      | 2926923              |      | 2926923  |      | 2926923  |      | 2926923  |      |
| R2                          | 0.03                   |      | 0.07     |      | 0.10     |      | 0.11     |      | 0.08                 |      | 0.09     |      | 0.09     |      | 0.10     |      |
| Wald tests of controls      |                        |      |          |      |          |      |          |      |                      |      |          |      |          |      |          |      |
| F                           |                        |      |          |      | 19102.01 |      | 28046.91 |      |                      |      |          |      | 9198.56  |      | 2974.44  |      |
| d.f.                        |                        |      |          |      | 2.00     |      | 5.00     |      |                      |      |          |      | 1.00     |      | 4.00     |      |
| p-value                     |                        |      |          |      | 0.00     |      | 0.00     |      |                      |      |          |      | 0.00     |      | 0.00     |      |

*Notes:* Estimates are from ordinary-least-squares regressions. Robust standard errors are shown in parentheses. Models 1 and 5 are the baseline models. Models 2 and 6 add subfield fixed effects. Models 3-4 (papers) and 7-8 (patents) add the field × year- and paper/patent-level controls, including adjustments for field × year level—*Number of new papers/patents*, *Mean number of papers/patents cited*, *Mean number of authors/inventors per paper/patent*—and paper/patent-level—*Number of papers/patents cited*, *Number of unlinked references*—characteristics. The reference categories for the year indicators are 1945 and 1980 for papers and patents, respectively. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. The Wald tests, shown below the models including the control variables, test the null hypothesis that parameters for the control variables are simultaneously zero using a two-sided F-test. Rejection of the null hypothesis is evidence that the control variables add to the explanatory power of the model.

\*p<0.1 \*\*p<0.05; \*\*\*p<0.01

**Supplementary Table 2:** Regression models of disruptiveness and the growth of knowledge

|                                                                | Sample: Patents View  |                       |                       | Sample: Web of Science |                        |                        |
|----------------------------------------------------------------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|------------------------|
|                                                                | (1)                   | (2)                   | (3)                   | (4)                    | (5)                    | (6)                    |
| Number of new works in the field during focal year (logged)    | 0.0036***<br>(0.0004) |                       |                       | -0.0042***<br>(0.0001) |                        |                        |
| Number of new works in the field during past 5 years (logged)  |                       | 0.0026***<br>(0.0002) |                       |                        | -0.0019***<br>(0.0001) |                        |
| Number of new works in the field during past 10 years (logged) |                       |                       | 0.0033***<br>(0.0002) |                        |                        | -0.0019***<br>(0.0001) |
| Constant                                                       | 0.0291***<br>(0.0030) | 0.0064<br>(0.0044)    | -0.0095*<br>(0.0051)  | 0.0519***<br>(0.0016)  | 0.0479**<br>(0.0016)   | 0.0491***<br>(0.0017)  |
| Year fixed effects                                             | Yes                   | Yes                   | Yes                   | Yes                    | Yes                    | Yes                    |
| Field fixed effects                                            | Yes                   | Yes                   | Yes                   | Yes                    | Yes                    | Yes                    |
| N                                                              | 3434055               | 3434055               | 3434055               | 2.16e+07               | 2.16e+07               | 2.16e+07               |
| R2                                                             | 0.06                  | 0.06                  | 0.06                  | 0.02                   | 0.02                   | 0.02                   |

*Notes:* This table evaluates the relationship between the number of new works (papers or patents, a proxy for the growth of knowledge) and  $CD_5$ . Estimates are from ordinary-least-squares regressions. Each coefficient is tested against the null hypothesis of being equal to 0 using a two-sided t-test. We do not adjust for multiple hypothesis testing. Robust standard errors are shown in parentheses.

\*p<0.1 \*\*p<0.05; \*\*\*p<0.01

**Supplementary Table 3:** Summary statistics for variables used in regression analyses  
WoS/Papers

| Variable                                            | <i>N</i> | <i>N<sub>unique</sub></i> | Mean  | SD    | Min   | Max     | Level of measurement |
|-----------------------------------------------------|----------|---------------------------|-------|-------|-------|---------|----------------------|
| Use of existing knowledge                           |          |                           |       |       |       |         |                      |
| Diversity of work cited                             | 24615975 | 8811                      | 0.98  | 0.01  | 0.88  | 1.00    | Field × year         |
| Ratio of self-citations to total work cited         | 21766010 | 5779                      | 0.16  | 0.20  | 0.00  | 1.00    | Publication          |
| Mean age of work cited                              | 21766010 | 74735                     | 9.22  | 5.55  | 0.00  | 110.00  | Publication          |
| Dispersion in age of work cited                     | 21766010 | 2511142                   | 6.32  | 4.22  | 0.00  | 53.00   | Publication          |
| Controls                                            |          |                           |       |       |       |         |                      |
| Mean age of team members                            | 24621907 | 22636                     | 13.67 | 10.76 | 0.00  | 80.00   | Publication          |
| Mean number of prior works produced by team members | 24656790 | 58843                     | 35.70 | 57.23 | 0.00  | 2381.00 | Publication          |
| Outcomes                                            |          |                           |       |       |       |         |                      |
| CD <sub>5</sub> index                               | 22479429 | 848779                    | 0.04  | 0.22  | -1.00 | 1.00    | Publication          |
| Fixed effects                                       |          |                           |       |       |       |         |                      |
| Publication year                                    | 24659076 | 66                        | –     | –     | –     | –       | Year                 |
| WoS extended subject                                | 24622172 | 151                       | –     | –     | –     | –       | Field                |

Patents View/Patents

| Variable                                            | <i>N</i> | <i>N<sub>unique</sub></i> | Mean  | SD    | Min    | Max     | Level of measurement |
|-----------------------------------------------------|----------|---------------------------|-------|-------|--------|---------|----------------------|
| Use of existing knowledge                           |          |                           |       |       |        |         |                      |
| Diversity of work cited                             | 3912211  | 1278                      | 0.98  | 0.01  | 0.84   | 1.00    | Field × year         |
| Ratio of self-citations to total work cited         | 3486663  | 6122                      | 0.07  | 0.18  | 0.00   | 1.00    | Publication          |
| Mean age of work cited                              | 3486663  | 47948                     | 7.67  | 4.15  | -36.00 | 34.00   | Publication          |
| Dispersion in age of work cited                     | 3486663  | 278012                    | 3.35  | 2.49  | 0.00   | 19.00   | Publication          |
| Controls                                            |          |                           |       |       |        |         |                      |
| Mean age of team members                            | 3911607  | 2116                      | 5.60  | 5.93  | 0.00   | 34.00   | Publication          |
| Mean number of prior works produced by team members | 3911607  | 8340                      | 10.51 | 49.70 | 0.00   | 2852.00 | Publication          |
| Outcomes                                            |          |                           |       |       |        |         |                      |
| CD <sub>5</sub> index                               | 3682052  | 58722                     | 0.12  | 0.30  | -1.00  | 1.00    | Publication          |
| Fixed effects                                       |          |                           |       |       |        |         |                      |
| Publication year                                    | 3912353  | 35                        | –     | –     | –      | –       | Year                 |
| NBER technology subcategory                         | 3912281  | 39                        | –     | –     | –      | –       | Field                |

## Supplementary references

1. Michel, J.-B. *et al.* Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
2. Milojević, S. Quantifying the cognitive extent of science. *Journal of Informetrics* **9**, 962–973 (2015).