Papers and patents are becoming less disruptive over time

https://doi.org/10.1038/s41586-022-05543-x

Received: 14 February 2022

Accepted: 8 November 2022

Published online: 4 January 2023

Check for updates

Michael Park[1], Erin Leahey[2] & Russell J. Funk[1]✉

Theories of scientific and technological change view discovery and invention as endogenous processes[1,2], wherein previous accumulated knowledge enables future progress by allowing researchers to, in Newton's words, 'stand on the shoulders of giants'[3–7]. 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[8,9]. Yet contrary to this view, studies suggest that progress is slowing in several major fields[10,11]. Here, we analyse these claims at scale across six decades, using data on 45 million papers and 3.9 million patents from six large-scale datasets, together with a new quantitative metric—the CD index[12]—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[1,13–17]. Subsequently, we link this decline in disruptiveness to a narrowing in the use of previous 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.

Although the past century witnessed an unprecedented expansion of scientific and technological knowledge, there are concerns that innovative activity is slowing[18–20]. Studies document declining research productivity in semiconductors, pharmaceuticals and other fields[10,11]. Papers, patents and even grant applications have become less novel relative to prior work and less likely to connect disparate areas of knowledge, both of which are precursors of innovation[21,22]. The gap between the year of discovery and the awarding of a Nobel Prize has also increased[23,24], suggesting that today's contributions do not measure up to the past. These trends have attracted increasing attention from policymakers, as they pose substantial threats to economic growth, human health and wellbeing, and national security, along with global efforts to combat grand challenges such as climate change[25,26].

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[19,27]. 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[18,28]. 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 previous 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'[3]. Moreover, to date, the evidence pointing to a slowdown is based on studies of particular fields, using disparate and domain-specific metrics[10,11], 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 mask differences in the degree to which individual works push the frontier.

We address these gaps in understanding by analysing 25 million papers (1945–2010) in the Web of Science (WoS) (Methods) and 3.9 million patents (1976–2010) in the United States Patent and Trademark Office's (USPTO) Patents View database (Methods). 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 four additional datasets—JSTOR, the American Physical Society corpus, Microsoft Academic Graph and PubMed—encompassing 20 million papers. Using these data, we join a new citation-based measure[12] 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[2,29,30], which distinguish between two types of breakthroughs. First, some contributions improve existing streams of knowledge, and therefore consolidate the status

[1]Carlson School of Management, University of Minnesota, Minneapolis, MN, USA. [2]School of Sociology, University of Arizona, Tucson, AZ, USA. ✉e-mail: rfunk@umn.edu

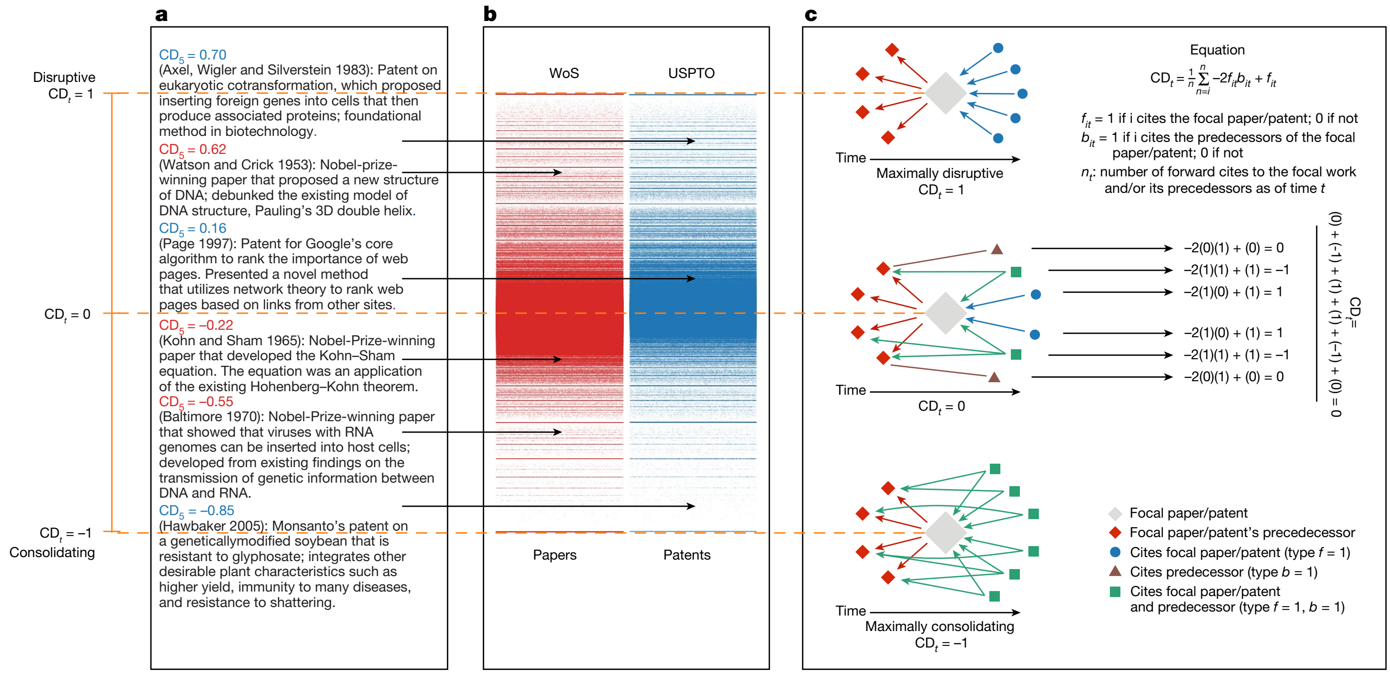


Fig. 1 | Overview of the measurement approach. This figure shows a schematic visualization of the CD index. **a**, CD index value of three Nobel Prize-winning papers[31,32,58] and three notable patents[59–61] in our sample, measured as of five years post-publication (indicated by CD5). **b**, Distribution of CD_5 for papers from WoS (n = 24,659,076) between 1945 and 2010 and patents from Patents View (n = 3,912,353) between 1976 and 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 helps to minimize overlapping points. Darker areas on each strip plot indicate denser regions of the distribution (that is, more commonly observed CD_5 values). Additional details on the distribution of the CD index are given in Extended Data Fig. 1. **c**, Three hypothetical citation networks, where the CD index is at 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.

quo. Kohn and Sham (1965)[31], a Nobel-winning paper used established theorems to develop a method for calculating the structure of electrons, which cemented the value of previous research. Second, some contributions disrupt existing knowledge, rendering it obsolete, and propelling science and technology in new directions. Watson and Crick (1953)[32], also a Nobel winner, introduced a model of the structure of DNA that superseded previous approaches (for example, Pauling's triple helix). Kohn and Sham and Watson and Crick were both important, but their implications for scientific and technological change were different.

We quantify this distinction using a measure—the CD index[12]—that characterizes the consolidating or disruptive nature of science and technology (Fig. 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 (for example, 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 (for example, the theorems Kohn and 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 Fig. 1 for the distribution of CD_5 among papers and patents and Extended Data Fig. 2 for analyses using alternative windows)[33]. For example, Watson and Crick and Kohn and Sham both received over a hundred citations within five years of being published. However, the Kohn and Sham paper has a CD_5 of −0.22 (indicating consolidation), whereas the Watson and Crick paper has a CD_5 of 0.62 (indicating disruption). The CD index has been validated extensively in previous research, including through correlation with expert assessments[12,34].

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 (Fig. 2a) and patents (Fig. 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 'social sciences') to 100% (where the average CD_5 decreased from 0.36 in 1945 to 0 in 2010 for '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 'life sciences and biomedicine' and physical sciences, and most marked and persistent in 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[19,27], which would probably predict greater heterogeneity in the decline, as it seems unlikely fields would 'consume' their low-hanging fruit at similar rates or 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 (for example, words used to create a new paradigm might differ from those that are used to develop an existing paradigm)[35,36].

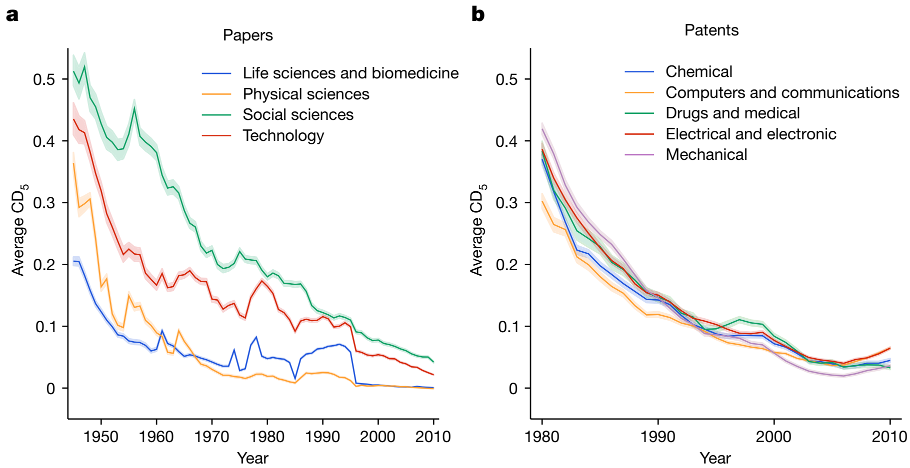


Fig. 2 | Decline of disruptive science and technology. a,**b**, Decline in CD_5 over time, separately for papers (**a**, n = 24,659,076) and patents (**b**, n = 3,912,353). For papers, lines correspond to WoS research areas; from 1945 to 2010 the magnitude of decline ranges from 91.9% (social sciences) to 100% (physical sciences). For patents, lines correspond to National Bureau of Economic Research (NBER) technology categories; from 1980 to 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, this pattern of decline is robust to adjustment for confounding from changes in publication, citation and authorship practices over time.

Therefore, if disruptiveness is declining, we would expect a decline in the diversity of words used in science and technology. To evaluate this, Fig. 3a,d documents the type-token ratio (that is, unique/total words) of paper and patent titles over time (Supplementary Information section 1). We observe substantial declines, especially in the earlier periods, before 1970 for papers and 1990 for patents. For paper titles (Fig. 3a), the decrease (1945–2010) ranges from 76.5% (social sciences) to 88% (technology); for patent titles (Fig. 3d), the decrease (1980–2010) ranges from 32.5% (chemical) to 81% (computers and communications). For paper abstracts (Extended Data Fig. 3a), the decrease (1992–2010) ranges from 23.1% (life sciences and biomedicine) to 38.9% (social sciences); for patent abstracts (Extended Data Fig. 3b), the decrease (1980–2010) ranges from 21.5% (mechanical) to 73.2% (computers and communications). In Fig. 3b,e, 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 previous work. Consistent with these trends in language, we also observe declining novelty in the combinations of previous work cited by papers and patents, based on a previously established measure of 'atypical combinations'[14] (Extended Data Fig. 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[35,36]. Figure 3 shows the most common verbs in paper (Fig. 3c) and patent titles (Fig. 3f) in the first and last decade of each sample (Supplementary Information section 2). Although 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 (for example, 'produce', 'form', 'prepare' and 'make'), discovery (for example, 'determine' and 'report') and perception (for example, '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 (for example, 'improve', 'enhance' and 'increase'), application (for example, 'use' and 'include') or assessment (for example, 'associate', 'mediate' and 'relate') of existing scientific and technological knowledge and artefacts. 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 (Methods and Fig. 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 (Extended Data Fig. 5, inset). These results suggest that the persistence of major breakthroughs—for example, measurement of gravity waves and COVID-19 vaccines—is not inconsistent with slowing innovative activity. In short, declining aggregate disruptiveness does not preclude individual 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 Fig. 5 shows that the decline in disruptiveness is unlikely to be due to other field-specific factors by decomposing variation in CD_5 attributable to field, author and year effects (Methods).

Declining rates of disruptive activity are unlikely to be caused by the diminishing quality of science and technology[22,37]. If they were, then the patterns seen in Fig. 2 should be less visible in high-quality work. However, when we restrict our sample to articles published in premier publication venues such as *Nature*, *Proceedings of the National Academy of Sciences* and *Science* or to Nobel-winning discoveries[38] (Fig. 5), 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

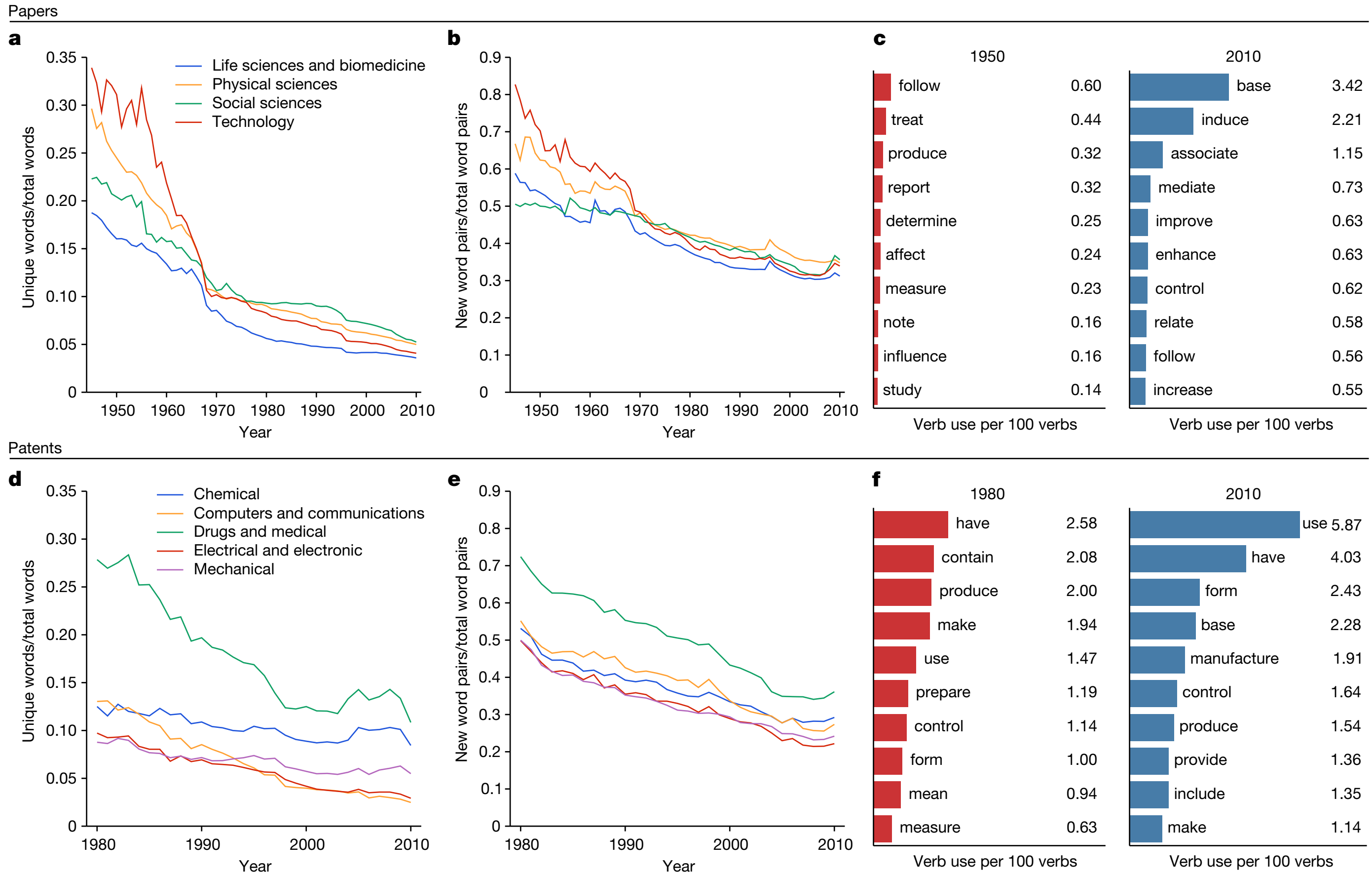


Fig. 3 | Decline of disruptive science and technology is visible in the changing language of papers and patents. **a**,**d**, Figures showing a decline in the diversity of language used in science and technology based on the unique/total words of paper titles from 1945 to 2010 (**a**, $n = 24{,}659{,}076$) and of patent titles from 1980 to 2010 (**d**, $n = 3{,}912{,}353$). **b**,**e**, Figures showing 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 to 2010 (**b**) and in Patents View patent titles from 1980 to 2010 (refs.[1,17]) (**e**). For papers in both **a** and **b**, lines correspond to WoS research areas ($n = 264$ WoS research area × year observations). For patents in both **d** and **e**, lines correspond to NBER technology categories ($n = 229$ NBER technology category × year observations). **c**,**f**, Figures showing 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.

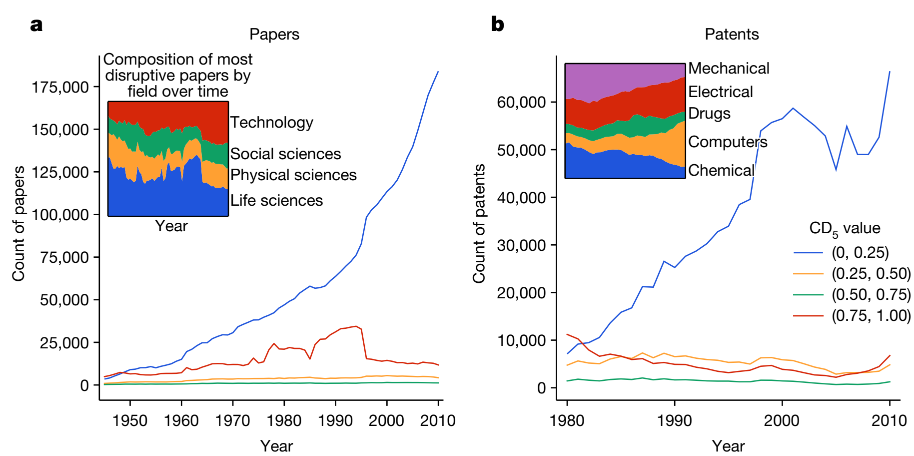
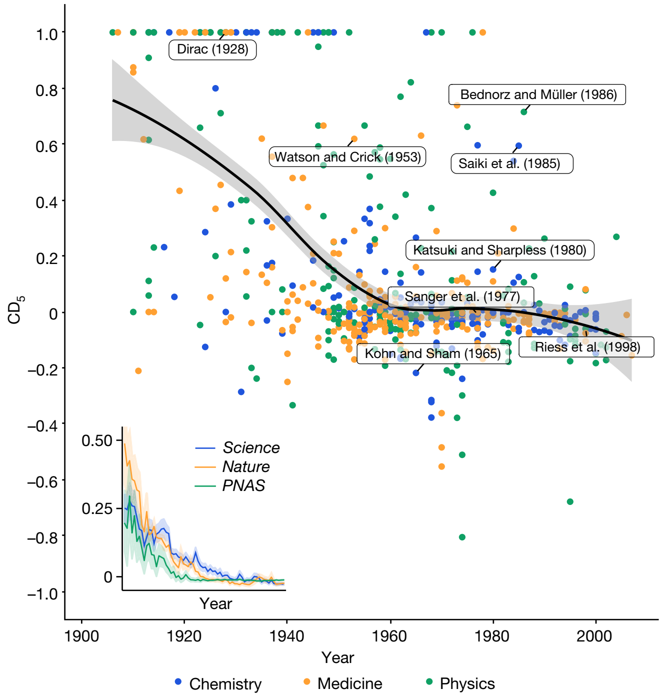


Fig. 4 | Conservation of highly disruptive work. This figure shows the number of disruptive papers (**a**, $n = 5{,}030{,}179$) and patents (**b**, $n = 1{,}476{,}004$) across four different ranges of CD_5 (papers and patents with CD_5 values in the range [−1.0, 0) are not represented in the figure). 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. 'Life sciences' denotes the life sciences and biomedicine research area; 'electrical' denotes the electrical and electronic technology category; 'drugs' denotes the drugs and medical technology category; and 'computers' denotes the computers and communications technology category.



Dirac (1928): Discovery of the Dirac equation of relativistic quantum theory
Watson and Crick (1953): Discovery of the structure of the DNA
Kohn and 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 and Sharpless (1980): Development of asymmetric epoxidation
Saiki et al. (1985): Discovery of polymerase chain reaction
Bednorz and Müller (1986): Discovery of superconductivity in cermaic materials
Riess et al. (1998): Discovery of the accelerating expansion of the universe

Fig. 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, $n = 635$), with several notable examples[31,32,58,62–66] highlighted. Colours indicate the three different journals in the inset plot; colours 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 WoS 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.

in JSTOR, the American Physical Society corpus, Microsoft Academic Graph and PubMed (Methods), the results of which are shown in Extended Data Fig. 6. We further show that the decline is not an artefact of the CD index by reporting similar patterns using alternative derivations[13,15] (Methods and Extended Data Fig. 7).

Declines in disruptiveness are also not attributable to changing publication, citation or authorship practices (Methods). First, using approaches from the bibliometrics literature[39–43], we computed several normalized versions of the CD index that adjusted for the increasing tendency for papers and patents to cite previous work[44,45]. Results using these alternative indicators (Extended Data Fig. 8a,d) were similar to those we reported in Fig. 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 × year level—number of new papers/patents, mean number of papers/patents cited, mean number of authors or inventors per paper—and paper or patent-level—number of papers or patents cited—factors. Predictions from these models indicated a decline in disruptive papers and patents (Extended Data Fig. 8b,e 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 behaviour, 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 Fig. 8c,f), 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 Fig. 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[3,6,7]. 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). We find a positive effect of the growth of knowledge on disruptiveness for papers, consistent with previous work[20]; 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 previous work[18,46], thereby limiting the 'effective' stock of knowledge. Using three proxies, we document a decline in the use of previous knowledge among scientists and inventors (Fig. 6). First, we see a decline in the diversity of work cited (Fig. 6a,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 (Fig. 6a (i),d(i)), which are also decreasing in semantic diversity (Fig. 6a (ii),d (ii)). Over time, scientists and inventors are increasingly citing the same previous work, and that previous work is becoming more topically similar. Second, we see an increase in self-citation (Fig. 6b,e), a common proxy for the continuation of one's pre-existing research stream[47–49], 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[50–52], is increasing (Fig. 6c,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 (Methods, Extended Data Table 1 and Supplementary Table 3), a pattern that holds even after accounting for the average age and number of previous 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 marked 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. Even though philosophers

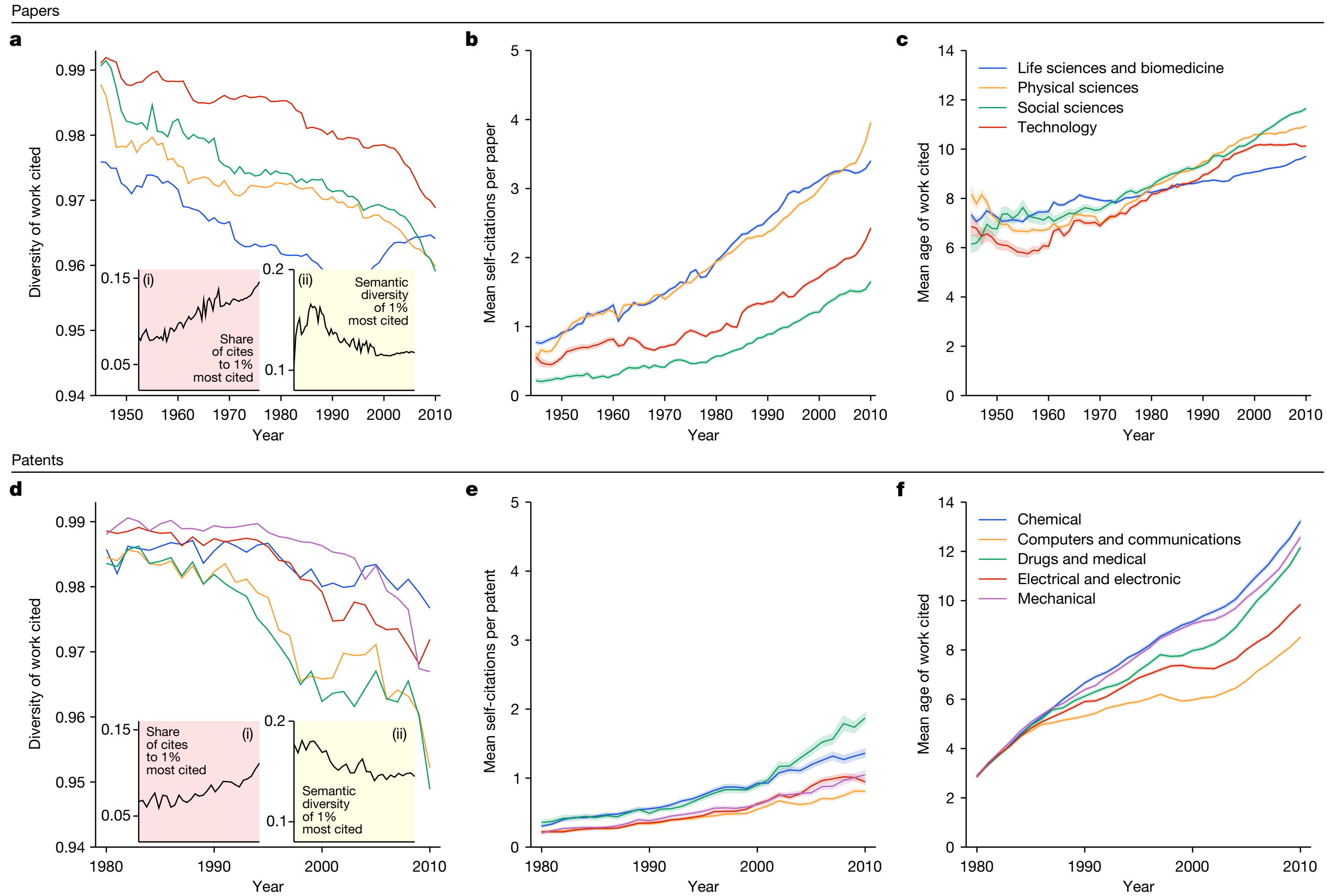


Fig. 6 | Papers and patents are using narrower portions of existing knowledge. **a**–**f**, Changes in the level of diversity of existing scientific and technological knowledge use among papers (**a**, $n = 264$ WoS research area × year observations; **b** and **c**, $n = 24{,}659{,}076$ papers) and patents (**d**, 229 NBER technology category × 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 (**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 (**a**(i) and **d**(i)) and in the semantic diversity of the top 1% most cited over time (**a**(ii) and **d**(ii)). 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 (that is, 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 (Methods and Extended Data Table 1).

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[53], but not scientific progress more generally.

Moreover, even though 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 suggests the potential importance of factors such as 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 suggests 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, even though research to date supports the validity of the CD index[12,34], it is a relatively new indicator of innovative activity and will benefit from future work on its behaviour and properties, especially across data sources and contexts. Studies that systematically examine the effect of different citation practices[54,55], 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[56], 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[57], 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.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05543-x.

1. Fleming, L. Recombinant uncertainty in technological search. *Manage. Sci.* **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. Popper, K. *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge, 2014).
5. Fleck, L. *Genesis and Development of a Scientific Fact* (Univ. Chicago Press, 2012).
6. Acemoglu, D., Akcigit, U. & Kerr, W. R. Innovation network. *Proc. Natl Acad. Sci. USA* **113**, 11483–11488 (2016).
7. Weitzman, M. L. Recombinant growth. *Q. J. Econ.* **113**, 331–360 (1998).
8. Tria, F., Loreto, V., Servedio, V. D. P. & Strogatz, S. H. The dynamics of correlated novelties. *Sci. Rep.* **4**, 1–8 (2014).
9. Fink, T. M. A., Reeves, M., Palma, R. & Farr, R. S. Serendipity and strategy in rapid innovation. *Nat. Commun.* **8**, 1–9 (2017).
10. Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* **10**, 428–438 (2011).
11. Bloom, N., Jones, C. I., Van Reenen, J. & Webb, M. Are ideas getting harder to find? *Am. Econ. Rev.* **110**, 1104–1144 (2020).
12. Funk, R. J. & Owen-Smith, J. A dynamic network measure of technological change. *Manage. Sci.* **63**, 791–817 (2017).
13. Bornmann, L., Devarakonda, S., Tekles, A. & Chacko, G. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quant. Sci. Stud.* **1**, 1242–1259 (2020).
14. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
15. Leydesdorff, L., Tekles, A. & Bornmann, L. A proposal to revise the disruption index. *Prof. Inf.* **30**, e300121 (2021).
16. Lu, C. et al. Analyzing linguistic complexity and scientific impact. *J. Informetr.* **13**, 817–829 (2019).
17. Hofstra, B. et al. The diversity–innovation paradox in science. *Proc. Natl Acad. Sci. USA* **117**, 9284–9291 (2020).
18. Jones, B. F. The burden of knowledge and the 'death of the renaissance man': is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).
19. Gordon, R. J. *The Rise and Fall of American Growth* (Princeton Univ. Press, 2016).
20. Chu, J. S. G. & Evans, J. A. Slowed canonical progress in large fields of science. *Proc. Natl Acad. Sci. USA* **118**, e2021636118 (2021).
21. Packalen, M. & Bhattacharya, J. NIH funding and the pursuit of edge science. *Proc. Natl Acad. Sci. USA* **117**, 12011–12016 (2020).
22. 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 Univ. Press, 2011).
23. Horgan, J. *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age* (Basic Books, 2015).
24. Collison, P. & Nielsen, M. *Science Is Getting Less Bang for its Buck* (Atlantic, 2018).
25. Nolan, A. Artificial intelligence and the future of science. *oecd.ai*, https://oecd.ai/en/wonk/ai-future-of-science (25 October 2021).
26. *Effective Policies to Foster High-risk/High-reward Research.* OECD Science, Technology, and Industry Policy Papers (OECD, 2021).
27. 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).
28. Einstein, A. *The World As I See It* (Citadel Press, 1949).
29. Arthur, W. B. The structure of invention. *Res. Policy* **36**, 274–287 (2007).
30. Tushman, M. L. & Anderson, P. Technological discontinuities and organizational environments. *Adm. Sci. Q.* **31**, 439–465 (1986).
31. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).
32. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
33. Bornmann, L. & Tekles, A. Disruption index depends on length of citation window. *Prof. Inf.* **28**, e280207 (2019).
34. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).
35. Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).
36. Brad Wray, K. Kuhn and the discovery of paradigms. *Philos. Soc. Sci.* **41**, 380–397 (2011).
37. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
38. Li, J., Yin, Y., Fortunato, S. & Wang, D. A dataset of publication records for Nobel laureates. *Sci. Data* **6**, 1–10 (2019).
39. Bornmann, L. & Marx, W. Methods for the generation of normalized citation impact scores in bibliometrics: which method best reflects the judgements of experts? *J. Informetr.* **9**, 408–418 (2015).
40. Waltman, L. A review of the literature on citation impact indicators. *J. Informetr.* **10**, 365–391 (2016).
41. Waltman, L. & van Eck, N. J. in *Springer Handbook of Science and Technology Indicators* (eds. Glänzel, W. et al.) 281–300 (Springer, 2019).
42. Bornmann, L. How can citation impact in bibliometrics be normalized? A new approach combining citing-side normalization and citation percentiles. *Quant. Sci. Stud.* **1**, 1553–1569 (2020).
43. Petersen, A. M., Pan, R. K., Pammolli, F. & Fortunato, S. Methods to account for citation inflation in research evaluation. *Res. Policy* **48**, 1855–1865 (2019).
44. Bornmann, L. & Mutz, R. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **66**, 2215–2222 (2015).
45. 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. *Humanit. Soc. Sci. Commun.* **8**, 1–15 (2021).
46. Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *Proc. Natl Acad. Sci. USA* **108**, 18910–18914 (2011).
47. Bonzi, S. & Snyder, H. Motivations for citation: a comparison of self citation and citation to others. *Scientometrics* **21**, 245–254 (1991).
48. Fowler, J. & Aksnes, D. Does self-citation pay? *Scientometrics* **72**, 427–437 (2007).
49. 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).
50. 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. *Sci. Adv.* **3**, e1601315 (2017).
51. Merton, R. K. Singletons and multiples in scientific discovery: a chapter in the sociology of science. *Proc. Am. Philos. Soc.* **105**, 470–486 (1961).
52. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
53. Leahey, E. Not by productivity alone: how visibility and specialization contribute to academic earnings. *Am. Sociol. Rev.* **72**, 533–561 (2007).
54. Tahamtan, I. & Bornmann, L. Core elements in the process of citing publications: conceptual overview of the literature. *J. Informetr.* **12**, 203–216 (2018).
55. 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).
56. Bhattacharya, J. & Packalen, M. *Stagnation and Scientific Incentives (Working Paper 26752)*, https://www.nber.org/papers/w26752 (2020).
57. Azoulay, P., Graff Zivin, J. S. & Manso, G. Incentives and creativity: evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011).
58. Baltimore, D. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209–1211 (1970).
59. Page, L. Method for node ranking in a linked database. US patent 6,285,999 (2001).
60. Axel, R., Wigler, M. H. & Silverstein, S. J. Processes for inserting DNA into eucaryotic cells and for producing proteinaceous materials. US patent 4,634,665 (1983).
61. Hawbaker, M. S. Soybean variety SE90346. US patent 6,958,436 (2005).
62. Katsuki, T. & Sharpless, K. B. The first practical method for asymmetric epoxidation. *J. Am. Chem. Soc.* **102**, 5974–5976 (1980).
63. Riess, A. G., et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009 (1998).
64. Dirac, P. A. M. The quantum theory of the electron. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **117**, 610–624 (1928).
65. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
66. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B Condens. Matter* **64**, 189–193 (1986).

Methods

WoS data

We limit our focus to research papers published between 1945 and 2010. Although the WoS data begin in the year 1900, the scale and social organization of science shifted markedly in the post-war era, thereby making comparisons with the present difficult and potentially misleading[67–69]. 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 archive 65 million documents published in 28,968 journals between 1900 and 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 and 2017. After eliminating non-research documents (for example, book reviews and 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 are 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 include the titles and abstracts for 6.5 million patents granted between 1976 and 2017. Following previous work[12], 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 only by the evidence we report, but also by previous research from diverse methodological and disciplinary perspectives[10,11,18–24]. 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 whereas overall, discovery and invention may be less disruptive over time, the high-level view taken in previous work may mask considerable heterogeneity. Put differently, aggregate evidence of slowing progress does not preclude the possibility that some subset of discoveries and inventions is highly disruptive.

To evaluate this possibility, we plot the number of disruptive papers (Fig. 4a) and patents (Fig. 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, whereas the latter two correspond to those that are more strongly so (for example, where we may expect to see major breakthroughs such as 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. Notably, 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 (Fig. 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 or 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 who produce science and technology. For example, exogenous factors such as 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 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) × year. We enter field fixed effects using granular subfield indicators (that is, 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 (for example, the basic subject matter); year fixed effects capture all year-specific factors that do not vary by field or author (for example, the state of communication technology); author (or inventor) fixed effects capture all author-specific factors that do not vary by field or year (for example, 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[70].

Results of this analysis are shown in Extended Data Fig. 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 (that is, 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 entered the field in more recent years may face a higher burden of knowledge and thus resort to building on narrower slices of existing work (for example, because of 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; although 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 artefacts of our choice of data sources. Although 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 such as changes in coverage (for example, 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 to 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 (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). Owing 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 Fig. 6, the downward trend in disruptiveness is evident across all samples.

Alternative bibliometric measures

Several recent papers have introduced alternative specifications of the CD index[12]. 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[13]. Bornmann et al.[13] 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[15]. Leydesdorff et al.[15] 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 Fig. 7a (papers) and b (patents). The blue lines indicate disruption based on Bornmann et al.[13] and the orange lines indicate disruption based on Leydesdorff et al.[15]. 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 artefact 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 marked expansion in publishing and patenting over the period of our study. This expansion has naturally increased the amount of previous work that is relevant to current science and technology and therefore at risk of being cited, a pattern reflected in the marked increase in the average number of citations made by papers and patents (that is, papers and patents are citing more previous work than in previous eras)[44,45]. Recall that the CD index quantifies the degree to which future work cites a focal work together with its predecessors (that is, 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 previous 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 marked 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[39–43], 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 previous work[13]. Larger values of N_k lead to smaller values of the CD index. Consequently, marked 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 previous work (N_b). The intuition behind this adjustment is that when a focal paper or patent cites more previous 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, although not necessarily as a result of the focal paper or patent being less disruptive. In the second version, which we call 'field × 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 previous work, N_k is also likely to be larger, thereby leading to lower values of the CD index, although again not necessarily as a result of the focal paper or patent being less disruptive. In cases in which either N_b or N_b^{mean} exceed the value of N_k, we set N_k to 0 (that is, 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' (that is, normalization by correcting for the effect of differences in lengths of references lists)[40].

In Extended Data Fig. 8, we plot the average values of both normalized versions of the CD index over time, separately for papers (Extended Data Fig. 8a) and patents (Extended Data Fig. 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 to be driven by 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 previous 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 (WoS subject areas for papers and NBER technology subcategories for patents). Finally, in Models 3–4 and 7–8, we add control variables for several field × year level—number of new papers orpatents, mean number of papers or patents cited, mean number of authors or inventors per paper—and paper- or patent-level—number of papers or 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 (that is, 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 eight 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 (Fig. 2). In Extended Data Fig. 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[14,71–73], we considered whether our results may be an artefact 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 non-trivial values on common network measures (for example, clustering)[74–76]. During the period spanned by our study, the citation networks of science and technology experienced significant change, with marked increases in both the numbers of nodes (that is, papers or patents) and edges (that is, 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[75,77] 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 (for example, 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 artefact 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 (for example, A–B and C–D) are selected from the underlying citation network, after which the algorithm attempts to swap the two endpoints of the edges (for example, 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' (that is, citations from other papers or patents to the focal paper or patent) and 'out-degree' (that is, citations from the focal paper or patent to other papers or patents) separately. Furthermore, we also required that the age distribution of citing and cited papers or patents was identical in the original and rewired networks. Specifically, swaps were only retained when the publication year of the original and candidate citations was 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[14], we created ten rewired copies of the observed citation networks for both papers and patents. After creating these rewired citation networks, we then recomputed CD_5. Owing to the large scale of the WoS data, we base our analyses on a random subsample of ten 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 ten rewired citation networks. Positive z scores indicate that the observed CD_5 value is greater (that is, more disruptive) than would be expected by chance; negative z scores indicate that the observed values are lesser (that is, more consolidating).

The results of these analyses are shown in Extended Data Fig. 8, separately for papers (Extended Data Fig. 8c) and patents (Extended Data Fig. 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 although there have been marked 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 to be an artefact 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 previous knowledge using regression models, predicting CD_5 for individual papers and patents, based on three indicators of previous 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 × 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 (for example, global economic trends). Field fixed effects account for field-specific factors that do not change over time (for example, 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' (for example, 'biochemistry and molecular biology', 'biophysics', 'biotechnology and applied microbiology', 'cell biology', 'developmental biology', 'evolutionary biology' and 'microbiology' are extended subjects within the life sciences and biomedicine research area) and 38 NBER technology subcategories (for example, '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' (that is, '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 previous works produced by team members'. Although 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 previous work available for self-citing. Similarly, although 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[78,79]. For example, older scientists and inventors may be more familiar with or more attentive to older work, or may actively resist change[80]. 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. In addition, the average CD_5 of a paper is 0.04 whereas 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 previous 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 in which there is more use of diverse work, there is greater disruption. Holding all other variables at their means, the predicted CD_5 of papers and patents increases by 303.5% and 1.3%, respectively, when the diversity of work cited increases by 1 s.d. 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 decreases by 622.9% and 18.5%, respectively, with a 1 s.d. 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 increases by a striking 2,072.4% and 58.4%, respectively, when the mean age of work cited increases by 1 s.d. (about nine and eight 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 previous knowledge may contribute to the production of less disruptive science and technology.

Reporting summary

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

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 WoS. 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 WoS are not publicly available, and were used under licence 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. Source data are provided with this paper.

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 v.3.10.6 (pandas v.1.4.3, numpy v.1.23.1, matplotlib v.3.5.2, seaborn v.0.11.2, spacy v.2.2, jupyterlab v.3.4.4) to wrangle, analyse and visualize data and to conduct statistical analyses. We used MariaDB v.10.6.4 to wrangle data. We used R v.4.2.1 (ggplot2 v.3.36, ggrepel v.0.9.0) to visualize data. We used StataMP v.17.0 (reghdfe v.5.7.3) to conduct statistical analyses.

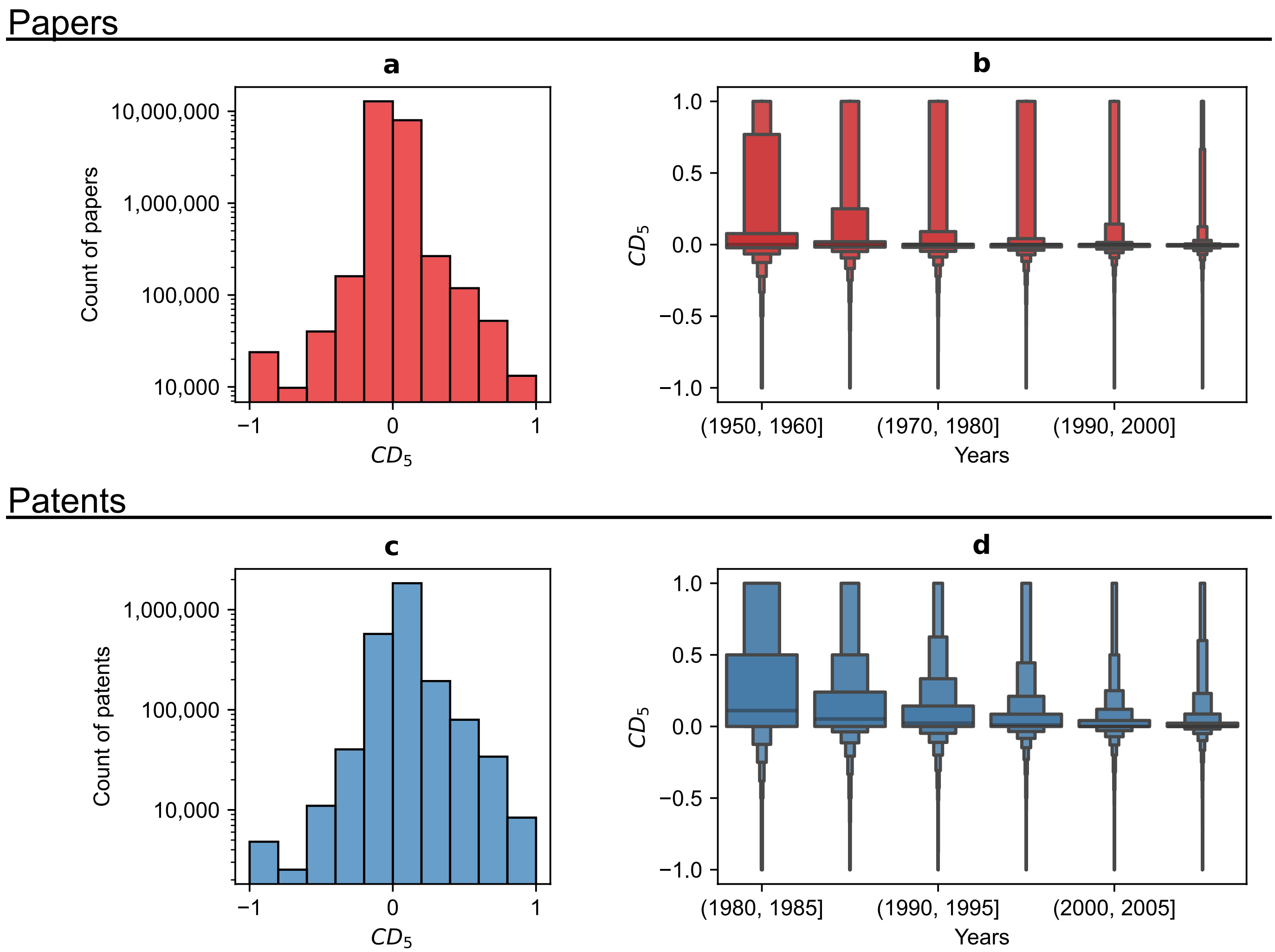
67. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
68. Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).
69. Jones, B. F., Wuchty, S. & Uzzi, B. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).
70. Grömping, U. Estimators of relative importance in linear regression based on variance decomposition. *Am. Stat.* **61**, 139–147 (2007).
71. Mukherjee, S., Uzzi, B., Jones, B. & Stringer, M. A new method for identifying recombinations of existing knowledge associated with high-impact innovation. *J. Prod. Innov. Manage.* **33**, 224–236 (2016).
72. Christianson, N. H., Sizemore Blevins, A. & Bassett, D. S. Architecture and evolution of semantic networks in mathematics texts. *Proc. R. Soc. A* **476**, 20190741 (2020).
73. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA* **98**, 404–409 (2001).
74. Newman, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).
75. Uzzi, B. & Spiro, J. Collaboration and creativity: the small world problem. *Am. J. Sociol.* **111**, 447–504 (2005).
76. Funk, R. J. Making the most of where you are: geography, networks, and innovation in organizations. *Acad. Manage. J.* **57**, 193–222 (2014).
77. Barabási, A.-L. *Network Science* (Cambridge Univ. Press, 2016).
78. Blau, D. M. & Weinberg, B. A. Why the US science and engineering workforce is aging rapidly. *Proc. Natl Acad. Sci. USA* **114**, 3879–3884 (2017).
79. Cui, H., Wu, L. & Evans, J. A. Aging scientists and slowed advance. Preprint at https://doi.org/10.48550/arXiv.2202.04044 (2022).
80. Azoulay, P., Fons-Rosen, C. & Graff Zivin, J. S. Does science advance one funeral at a time? *Am. Econ. Rev.* **109**, 2889–2920 (2019).

Acknowledgements This study was supported by the National Science Foundation (grant Nos. 1829168, 1932596 and 1829302).

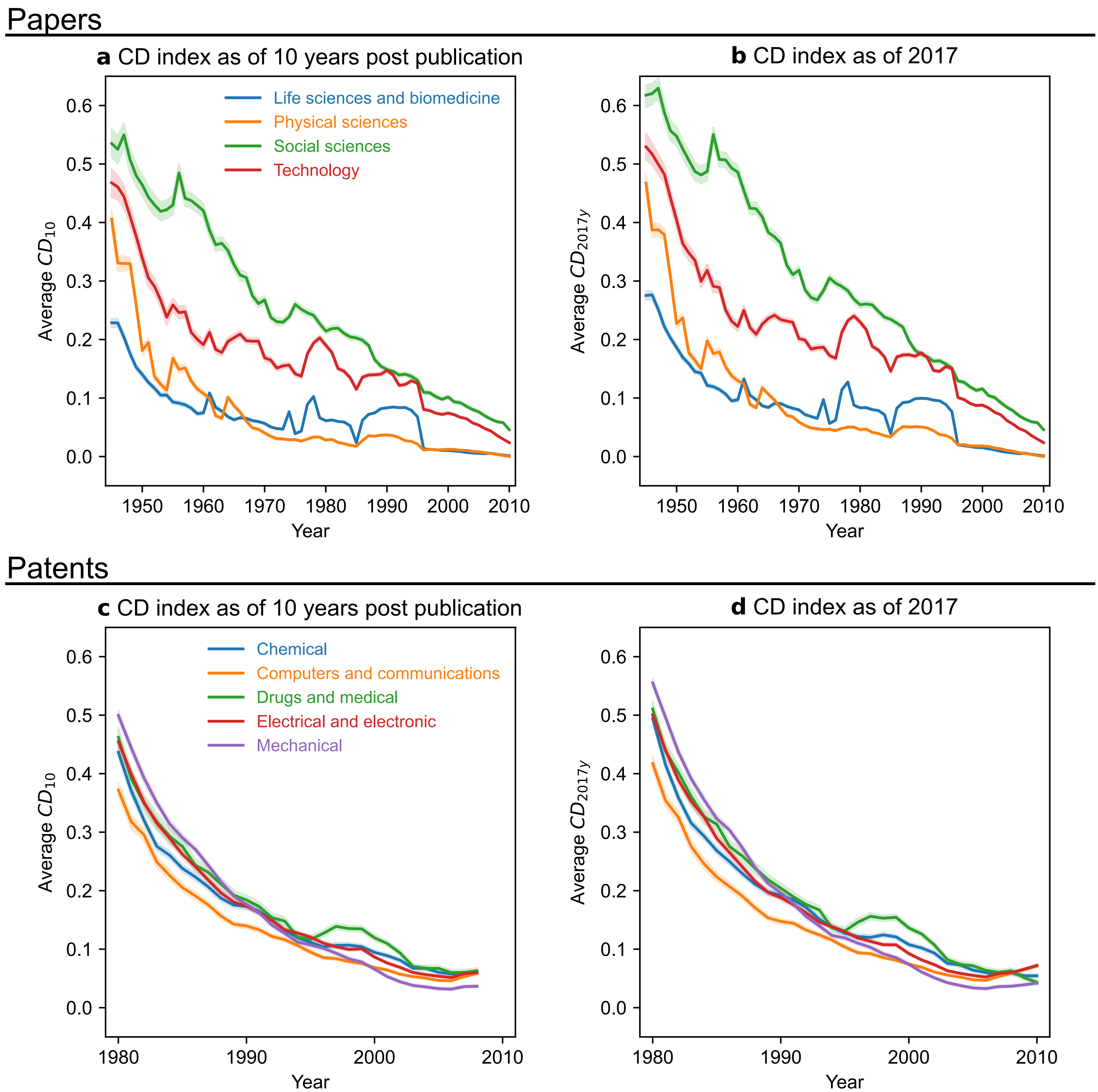
Author contributions R.J.F. and E.L. collaboratively contributed to the conception and design of the study. R.J.F. and M.P. collaboratively contributed to the acquisition, analysis and interpretation of the data. R.J.F. created software used in the study. R.J.F., E.L. and M.P. collaboratively drafted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

Competing interests The authors declare no competing interests.

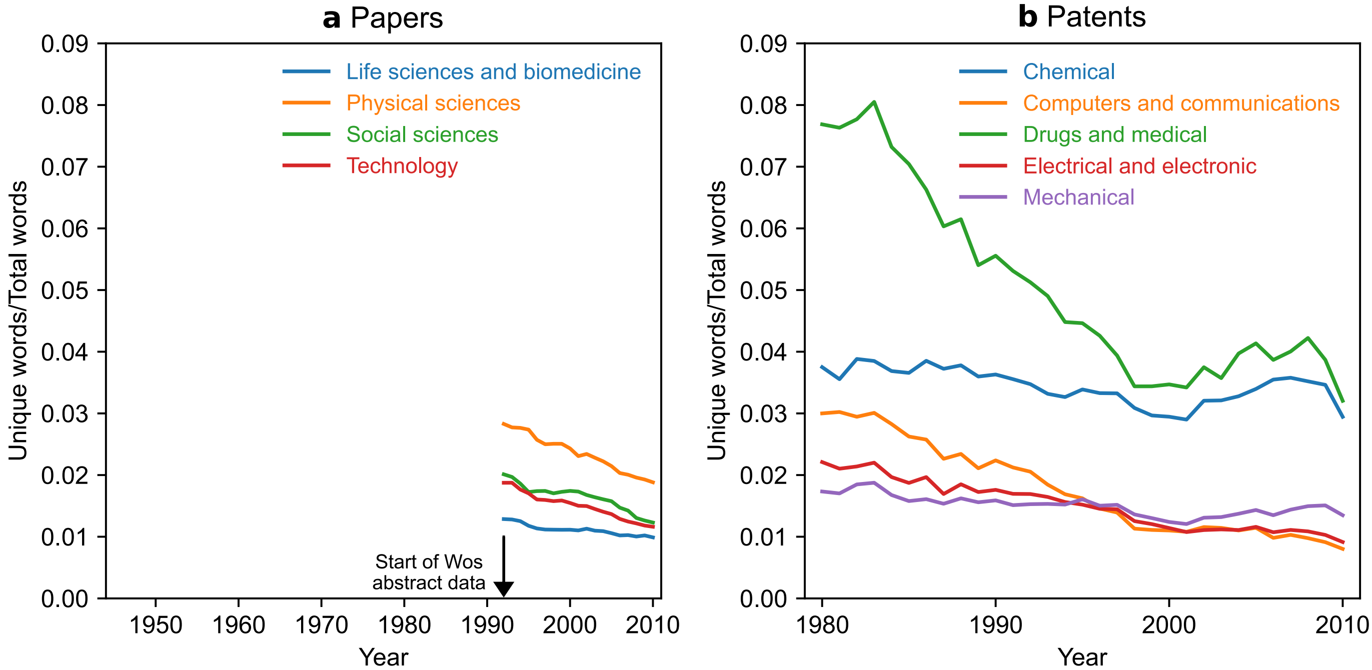
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05543-x.
Correspondence and requests for materials should be addressed to Russell J. Funk.
Peer review information *Nature* thanks Diana Hicks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



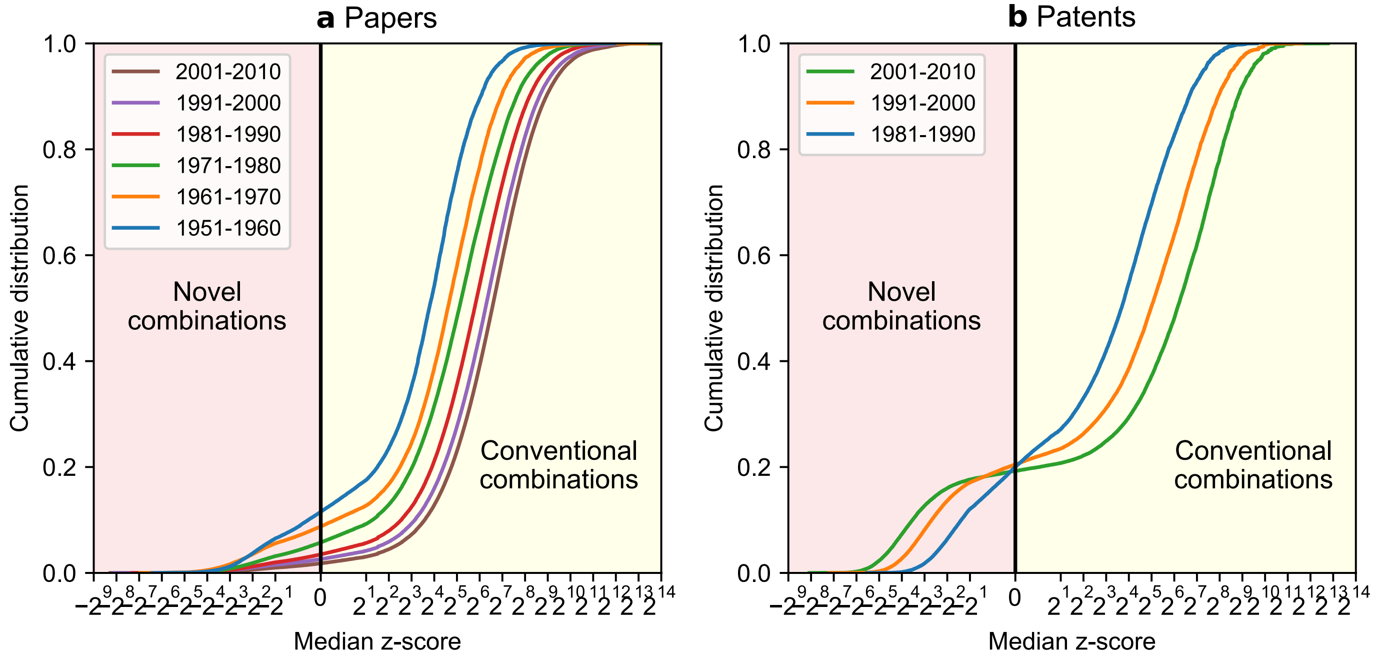
Extended Data Fig. 1 | Distribution of CD_5. This figure gives an overview of the distribution of CD_5 for papers ($n = 24,659,076$) and patents ($n = 3,912,353$). Panels **a** and **c** show counts of papers and patents over discrete intervals of CD_5. Panels **b** and **d** show the distribution of 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.



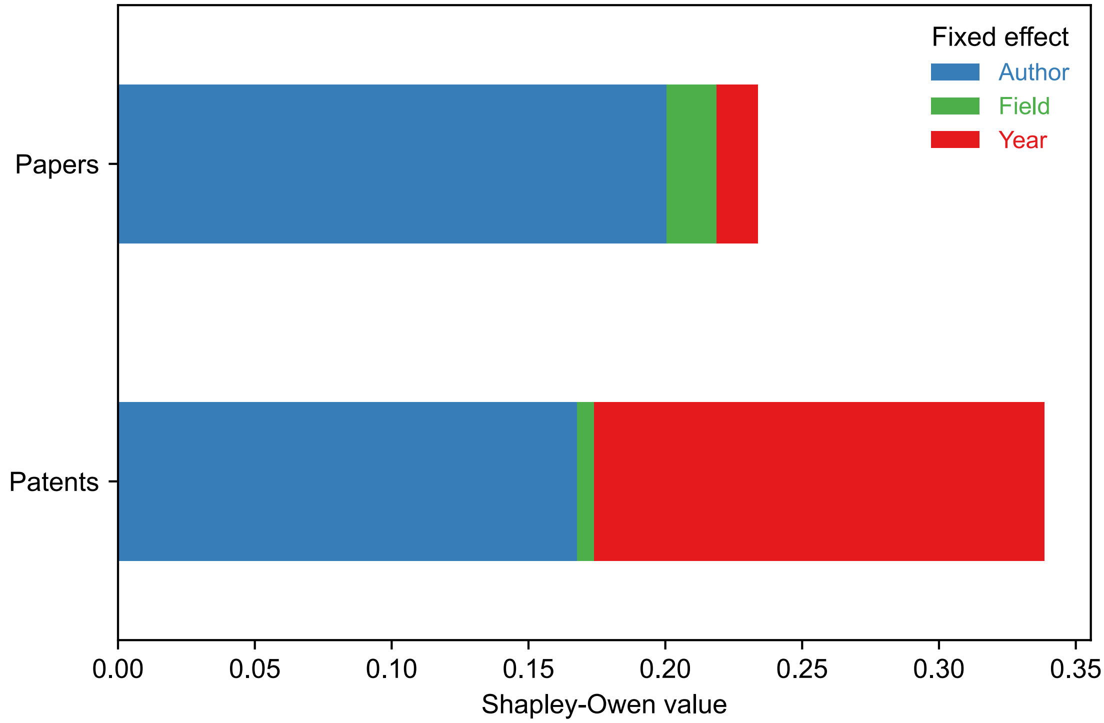
Extended Data Fig. 2 | CD index measured using alternative forward citation windows. This figure evaluates the sensitivity of our results to the use of different forward citation windows when computing the CD index for papers (n = 24,659,076) and patents (n = 3,912,353). In the main text, the index is 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.



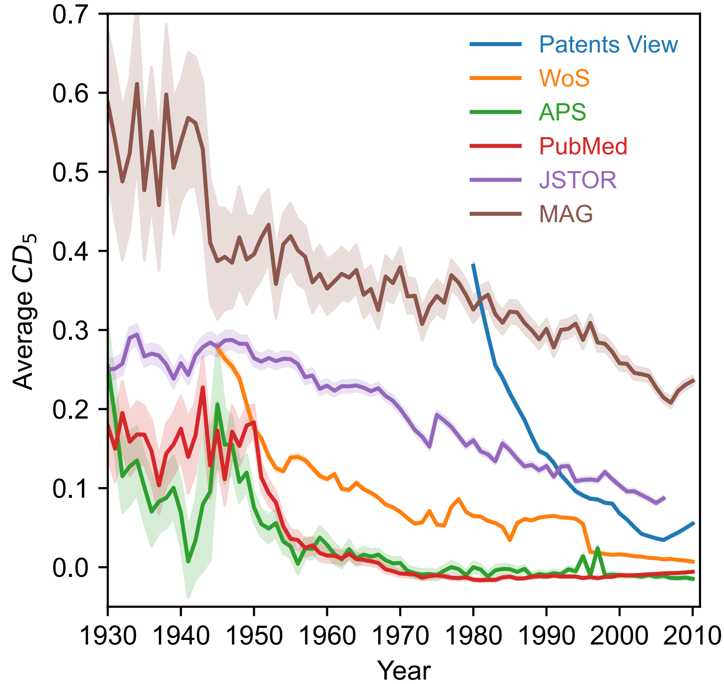
Extended Data Fig. 3 | Diversity of language use in science and technology over time. This figure shows changes in the ratio of unique to total words (also known as the type-token ratio) over time based on data from the abstracts of papers (**a**, n = 76 WoS research area × year observations) and patents (**b**, n = 229 NBER technology category × 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.



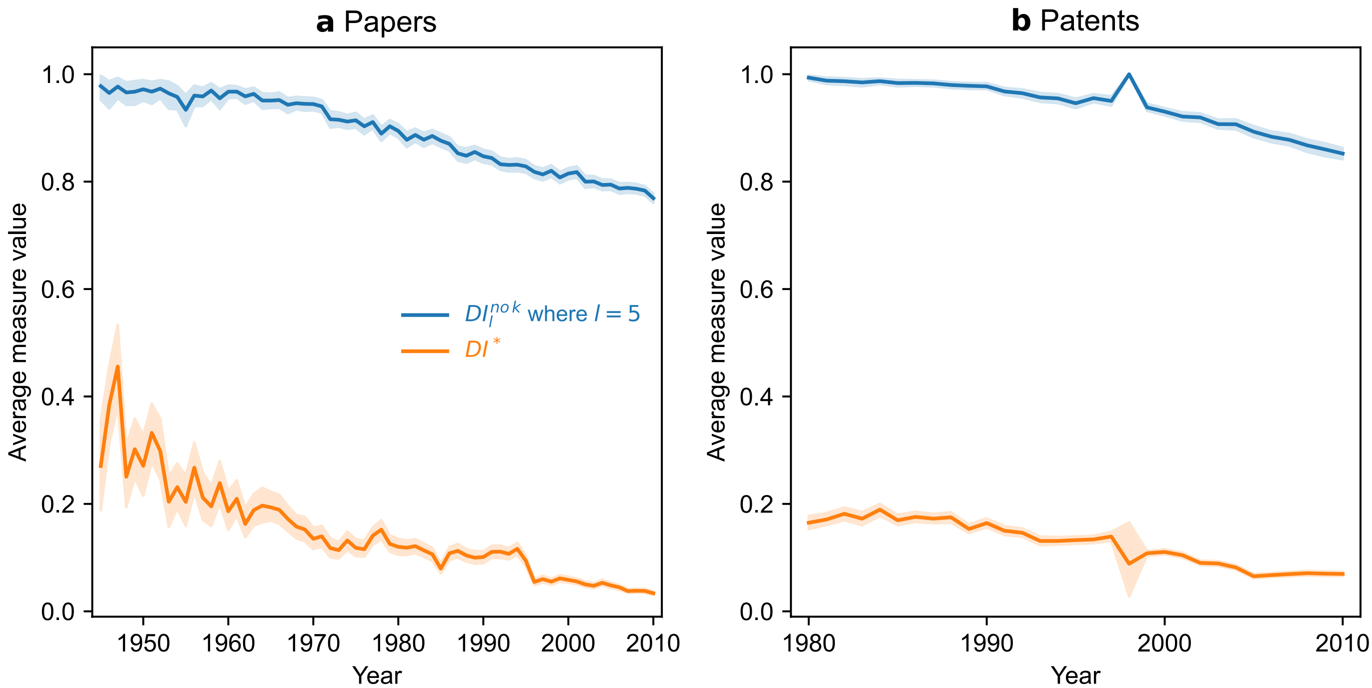
Extended Data Fig. 4 | Declining combinatorial novelty. This figure shows changing patterns in the combinatorial novelty/conventionality of papers (**a**, n = 24,659,076) and patents (**b**, n = 3,912,353), using a previously proposed measure of "atypical combinations"[14]. The measure quantifies the degree to which the prior work cited by a paper or patent would be expected by chance. For papers, we follow prior work[14] and consider combinations of cited journals. 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* × *Cell*, *Nature* × *Science*, and *Science* × *Cell*. To determine the degree to which each combination would be expected by chance, the frequency of observed pairings is 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[14], 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 rightward, indicating greater conventionality.



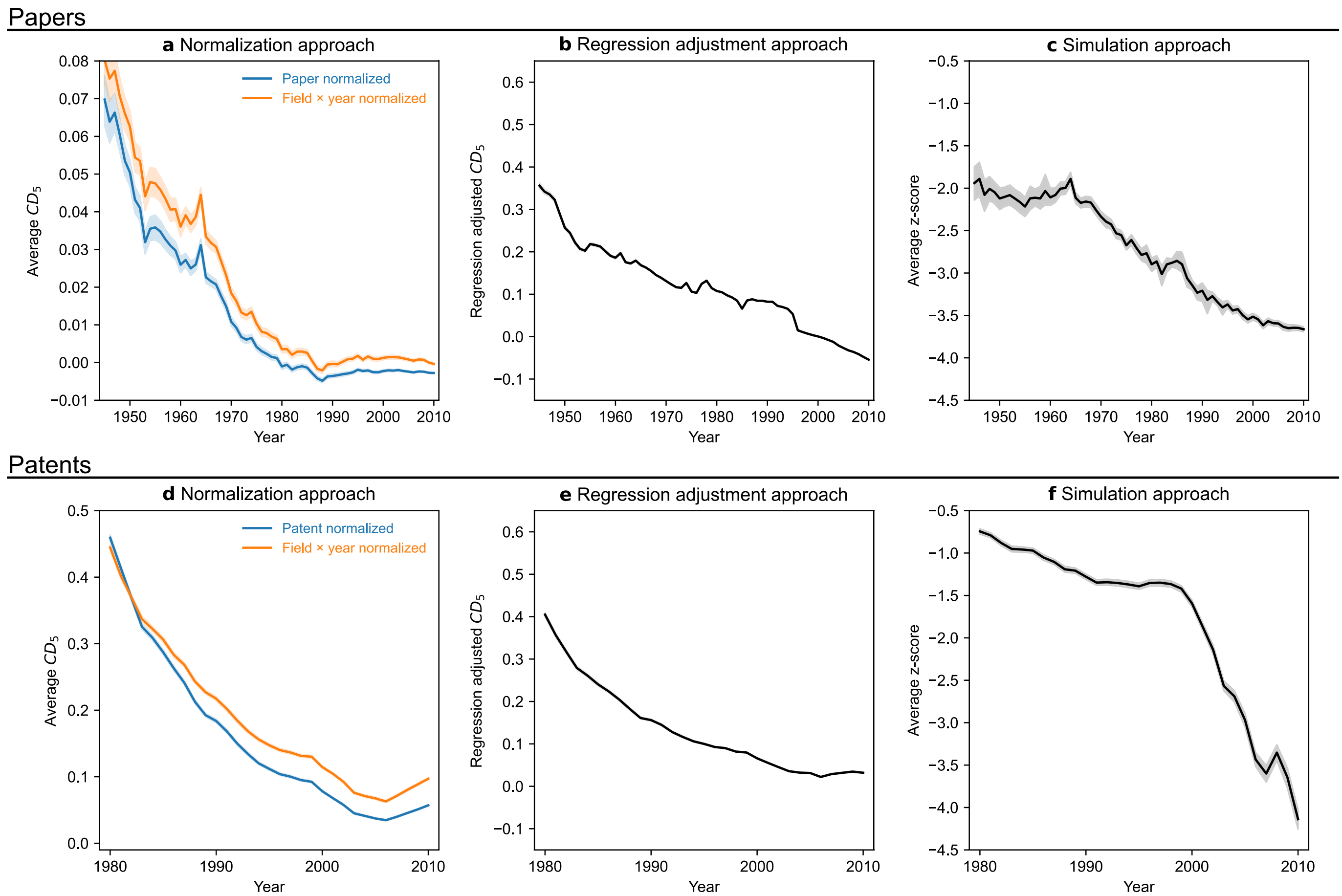
Extended Data Fig. 5 | Contribution of field, year, and author effects. This figure shows the relative contribution of field, year, and author fixed effects to the adjusted R^2 in regression models predicting CD_5. The top bar shows the results for papers (n = 80,607,091 paper × author observations);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



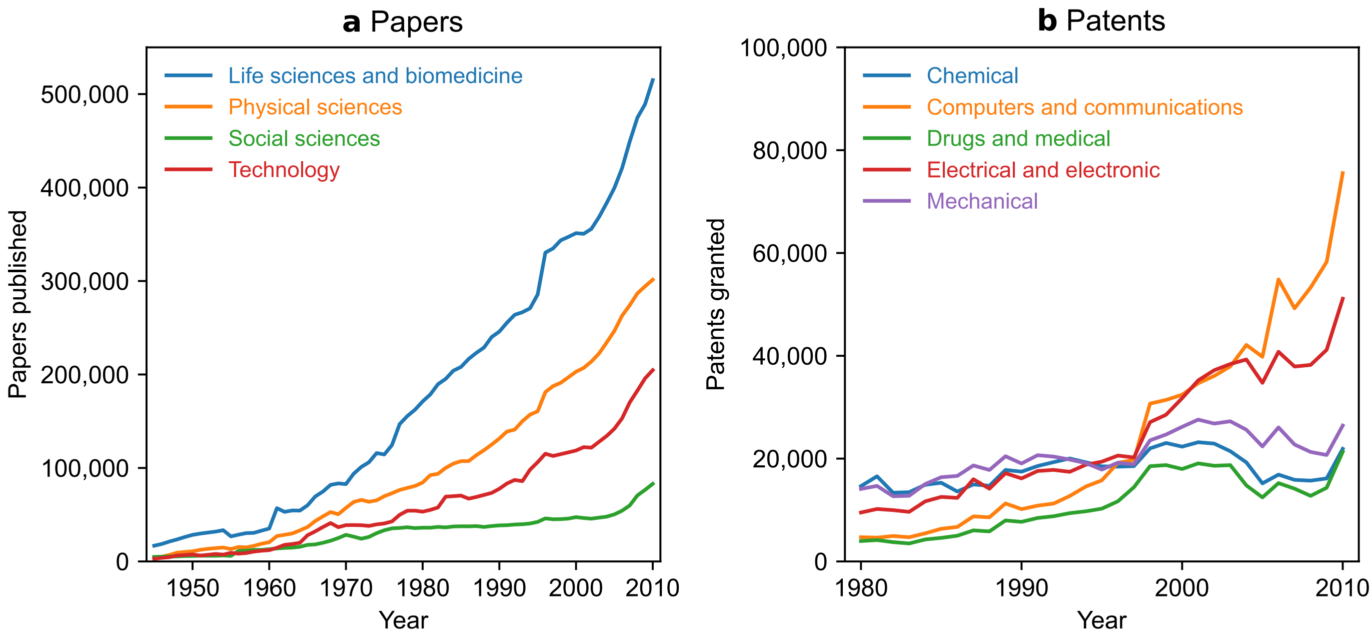
Extended Data Fig. 6 | CD index over time across data sources. This figure shows changes in CD_5 over time across four additional data sources (the WoS [n = 24,659,076] and Patents View [n = 3,912,353] lines are included for reference): JSTOR (n = 1,703,353), the American Physical Society corpus (n =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



Extended Data Fig. 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.[13], 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.[15], 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 artefact of our operationalization of disruption.



Extended Data Fig. 8 | Robustness to changes in publication, citation, and authorship practices. This figure evaluates whether 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/patent). Orange lines indicate normalization at the field and year level (accounting for the mean number of citations made by papers/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 × 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 year indicator 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 Fig. 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.

Extended Data Table 1 | Regression models of disruptiveness and the use of prior knowledge

	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***	0.3339***	0.1574***	0.4578***	0.1583***	0.1587***	0.1151***	0.1119***	0.0873***	1.3737***	0.0688***	0.0692***
	(0.0062)	(0.0062)	(0.0061)	(0.0025)	(0.0061)	(0.0061)	(0.0158)	(0.0158)	(0.0156)	(0.0080)	(0.0156)	(0.0156)
Ratio of self-citations to total work cited		-0.0191***	-0.0091***	-0.0118***	-0.0104***	-0.0107***		-0.0606***	-0.0557***	-0.0671***	-0.0585***	-0.0597***
		(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)		(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Mean age of work cited			0.0034***	0.0027***	0.0028***	0.0028***			0.0074***	0.0008***	0.0046***	0.0046***
			(0.0000)	(0.0000)	(0.0000)	(0.0000)			(0.0000)	(0.0001)	(0.0001)	(0.0001)
Dispersion in age of work cited			-0.0051***	-0.0069***	-0.0063***	-0.0063***			-0.0205***	-0.0370***	-0.0293***	-0.0293***
			(0.0000)	(0.0000)	(0.0000)	(0.0000)			(0.0001)	(0.0001)	(0.0001)	(0.0001)
Mean age of work cited × Dispersion in age of work cited				0.0001***	0.0001***	0.0001***				0.0013***	0.0009***	0.0009***
				(0.0000)	(0.0000)	(0.0000)				(0.0000)	(0.0000)	(0.0000)
Mean age of team members						0.0000***						0.0002***
						(0.0000)						(0.0000)
Mean number of prior works produced by team members						0.0000***						-0.0000***
						(0.0000)						(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

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.

nature portfolio

Corresponding author(s): Russell J Funk

Last updated by author(s): Oct 26, 2022

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	We used Python v3.10.6 and MariaDB v.10.6.4 to collect and organize raw XML and CSV files obtained from the relevant data publishers.
Data analysis	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.

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

Data

Policy information about availability of data

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

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

Data 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.

Field-specific reporting

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

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

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

Behavioural & social sciences study design

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

Study description	We chiefly rely on archival, observational, and quantitative citation and textual data on scientific and technological works (e.g., papers and patents), which were obtained from the following sources: Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. All analyses conducted were retrospective in nature. Our aim was to document the changing nature of innovative activity in science and technology as reflected in the data across time and to closely investigate the large-scale temporal patterns that emerged.
Research sample	There were no human subjects in our sample. We analyzed archival, observational, and quantitative data from Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. Our sample included approximately 45 million scientific papers and 3.9 million utility patents from 1945-2010. Among the papers, we analyzed their 390 million citations, 25 million titles, and 13 million abstracts. Among the patents, we analyzed their 35 million citations, 3.9 million titles, and 3.9 million abstracts. Our primary sample is composed of Web of Science and Patents View data. Web of Science data from the post-World War II era is among the most accurate, detailed, and comprehensive data available for studying the dynamics of science. Similarly, patent data from Patents View represents the population of machine-readable U.S. patent data available to the public.
Sampling strategy	There was no power analysis used to determine sample size. Rather, we analyzed the population of all available records (with complete data) from the following archival sources: Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed. This approach allowed us to analyze all available data to derive the most robust empirical results.
Data collection	We used Python v3.10.6 and MariaDB v.10.6.4 to initially collect and organize the raw XML and CSV data we obtained from the publishers of Web of Science, Patents View, JSTOR, the American Physical Society, Microsoft Academic Graph, and PubMed.
Timing	Data collection lasted about 30 days during the month indicated in parentheses for each of the dataset: Web of Science (December, 2018), Patents View (June, 2018), JSTOR (November, 2016), American Physical Society (August, 2019), Microsoft Academic Graph (February, 2021), and PubMed (July, 2019).
Data exclusions	N/A: No data was excluded. We analyzed all data that was made available to us from the publishers.
Non-participation	N/A: The study only conducted retrospective analysis of bibliometric data and had no participants (and thus no drop outs or declines).
Randomization	N/A: We used an observational study design on archival data and therefore there was no randomization.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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