

RESEARCH IMPACT

The dual frontier: Patented inventions and prior scientific advance

Mohammad Ahmadpoor^{1,2} and Benjamin F. Jones^{1,2,3*}

The extent to which scientific advances support marketplace inventions is largely unknown. We study 4.8 million U.S. patents and 32 million research articles to determine the minimum citation distance between patented inventions and prior scientific advances. We find that most cited research articles (80%) link forward to a future patent. Similarly, most patents (61%) link backward to a prior research article. Linked papers and patents typically stand 2 to 4 degrees distant from the other domain. Yet, advances directly along the patent-paper boundary are notably more impactful within their own domains. The distance metric further provides a typology of the fields, institutions, and individuals involved in science-to-technology linkages. Overall, the findings are consistent with theories that emphasize substantial and fruitful connections between patenting and prior scientific inquiry.

Scientific research can propel both fundamental understanding and practical application, but the extent to which scientific advances support technological progress is unclear (1–3). According to the “linear model” of science, basic research, focused on understanding, provides a foundation for eventual technological applications (1, 4–7). For example, Riemannian geometry, an abstract mathematical advance that was initially widely ignored, later proved essential to Einstein’s development of general relativity and, ultimately, to time dilation corrections in the Global Positioning System. In biology, basic research into extremophile bacteria later proved essential to the development of the polymerase chain reaction, the DNA amplification technique that is vital to modern biotechnology applications. Such examples illustrate the potential value of the linear model as a conception of scientific and technological progress, a view that helps motivate the public case for supporting scientific research (1, 8, 9).

At the same time, many observers argue that basic research rarely pays off in practical application or that practical advances typically proceed without any inspiration from basic research (10–14). These views suggest a potentially substantial disconnect between the knowledge outputs of public science institutions, such as research universities or government laboratories, and inventive outputs in the private sector. Other scholars argue for a richer interplay between scientific and technological progress. Characterizing scientific progress as advances in understanding and technological progress as advances in use, a common theme emphasizes that investigators focused on questions of use, engaged in solving real problems, may in turn generate new understandings and progress in basic science (2, 15–17).

For example, Pasteur’s germ theory of disease was closely intertwined with his work on industrial fermentation and food safety applications, and the development of the second law of thermodynamics was inspired by Carnot’s practical interest in the efficiency limits of steam engines (2, 7). In these cases, new understandings of nature are seen less as independent exercises of human curiosity that pay off in unexpected, future applications than as insights that spring up along the technological frontier.

Amid these diverse views of the interplay between scientific and technological progress, there are many anecdotes but little systematic evidence. Our starting point is an integrated citation network that traces references from all 4.8 million patents issued by the U.S. Patent and Trademark Office (USPTO) from 1976 to 2015 to all 32 million journal articles published from 1945 to 2013 as indexed by the Web of Science (WOS), the world’s largest collection of scientific research. The citation network begins by locating patents that directly cite journal articles, which defines a “paper-patent boundary” where practical inventions and scientific advances are linked (18–21). The network further determines the minimum citation distance for all other papers and patents to this boundary, creating a measure of distance that can be applied across a broad landscape of scientific and technological progress. We further integrate information about fields, individuals, and institutions (universities, government laboratories, and publicly traded firms) for each paper and patent. The supplementary materials detail the underlying data sources and further discuss the use of citation networks to measure knowledge flows, including patent-to-paper citations (22–26).

Figure 1A presents a schematic of the integrated citation network and introduces our metric. Formally, we define the distance metric $D_i \in \{1, 2, 3, \dots\}$ for each patent or paper i . When a patent directly cites a paper, both nodes receive $D_i = 1$, representing patents and papers at the “patent-paper boundary.” For the set of all other

paper and patents, we recursively determine the minimum citation distance to this boundary. Namely, a paper i with $D_i = n + 1$ is one that is cited by a paper j with $D_j = n$ and is not cited by any paper k with $D_k < n$. Similarly, a patent i with $D_i = n + 1$ is one that cites a patent j with $D_j = n$ and does not cite any patent k with $D_k < n$. Paper and patents that cannot be connected at any distance to the paper-patent boundary are described as “unconnected.” Note that the graph is directed: We trace citations backward in time, using the references in each patent and paper and jumping from the patent to the paper domain where $D_i = 1$.

Our first results concern connectivity, considering the extent to which papers or patents exist in independent spheres. As shown in Fig. 1B, the patent-paper citation network has been dominated by a single connected component. A majority of patents—60.5%—made references that could ultimately be traced to science and engineering papers. Similarly, among all science and engineering papers that received at least one citation, 79.7% could ultimately be connected to a patent. In short, we find majority connectivity, where the substantial majority of cited research articles can be linked to a future patent, and the modest majority of patents can be linked to prior scientific research.

At the boundary, 0.759 million patents directly cited 1.41 million papers, representing 21% of all connected patents and 10% of all connected papers (Fig. 1C). Although these numbers are substantial, the broader picture that emerges in Fig. 1C is one of indirect connectivity. The modal connected science and engineering paper was 3 degrees from the nearest patent. The modal connected patent was 2 degrees from the nearest paper. Looking between 2 and 4 degrees of the patent-paper boundary captures 68% of all connected patents and 79% of all connected papers.

Our second set of results applies the distance metric to characterize fields. We used 185 WOS field classifications for science and engineering papers and the 388 primary USPTO technology classes that contained at least 20 patents in the citation network. For each field or class, Fig. 2A presents the mean distance, D_{mean} , among connected papers or patents as well as the percentage connectivity (i.e., the percentage of papers or patents in that field for which D exists). Here we see the enormous variation across fields. D_{mean} ranged from 2.00 to 5.90 across science fields and from 1.17 to 5.65 across patent classes.

Examining patents, the technology classes closest to the paper-patent boundary include combinatorial chemistry, molecular biology, superconducting technology, and artificial intelligence, all of which had $D_{\text{mean}} < 1.50$. The most distant technology classes concern subjects such as locks, buttons, fasteners, envelopes, fire escapes, and chairs, all of which had $D_{\text{mean}} > 4.75$. To further characterize this variation, we examined the full D distributions for several major technology classes (Fig. 2B). For example, we see that $D_{\text{mode}} = 1$ for “multicellular living organism” patents, where 85% directly cited papers, whereas $D_{\text{mode}} = 5$ for

¹Northwestern University, Kellogg School of Management, Evanston, IL 60208, USA. ²Northwestern University Institute on Complex Systems and Data Science, Evanston, IL 60208, USA. ³NBER, Cambridge, MA 02138, USA.

*Corresponding author. Email: bjones@kellogg.northwestern.edu

“chairs and seats” patents, for which only 0.3% directly cited papers.

Examining papers in Fig. 2A, we see that mathematics proved the field most distant from the patent frontier ($D_{\text{mean}} = 4.97$). Meanwhile, the fields closest to the patent frontier include nanoscience and nanotechnology, materials science and biomaterials, and computer science hardware and architecture, all with $D_{\text{mean}} < 2.35$. Figure 2B provides the full D distributions for several major fields. Connected papers in mathematics, often considered a basic field of inquiry but one that can also be applied, had $D_{\text{mode}} = 4$ but with high variance. Astronomy and astrophysics also had $D_{\text{mode}} = 4$ but with a sharper peak and typically greater proximity to the patent-paper boundary. By contrast, biochemistry and molecular biology papers had $D_{\text{mode}} = 2$, and computer science papers had $D_{\text{mode}} = 1$, where 42% of connected computer science papers were directly cited by patents. This application to scientific fields suggests the potential usefulness of the distance metric for quantifying and tightening traditional but loose descriptors around “basic” and “applied” scientific research. The supplementary materials

show that the field ordering by distance to the patent-paper boundary is robust to different referencing tendencies across fields, to dropping patent-examiner citations in patents, and considers a null model (figs. S1, S8, and S9). Tables S1 and S2 provide the mean, mode, and standard deviation of the distance metric and percentage connectivity for all patent technology classes and all WOS fields.

Figure S2 considers a related concept of distance: time. We calculated the total time period, T_i , in years along the shortest citation path between a paper and a patent. This time period is the difference between the patent’s application year and the paper’s publication year. At the boundary, where $D = 1$, there was a mean delay of 6.66 years. By $D = 6$, the mean delay was 19.62 years for papers and 22.70 years for patents. Figure S2 further shows that the temporal distance varied substantially across fields, commensurate with the citation distance variation in Fig. 2A.

Figure 3 considers impact. A common measure of impact for a scientific paper or patent is the number of citations it receives, and a transparent, field-independent metric considers

the probability of a “home run,” defined as being in the upper 5% of citations received in that field and year (27–29). Figure 3A examines the probability of such home-run papers and patents. Patents that drew directly on scientific papers (i.e., $D = 1$ patents) were found to be unusually heavily cited by other patents, appearing as home runs 7.62% of the time, or 52.4% more often than the background rate. Other connected patents (i.e., $D \geq 2$ patents) were home runs at approximately the background rate. Figure S3 shows more generally that impact decayed smoothly with distance from the patent-paper frontier. Meanwhile, patents whose cited prior art was disconnected from the corpus of papers were home runs at a rate of 3.74%, or 25.2% less often than the background rate. Looking at papers in Fig. 3A, journal articles directly cited by a patent (i.e., $D = 1$ papers) were 3.72 times more likely to be highly cited by other papers. In other words, the patent-paper boundary appears populated by advances that are especially impactful within their own domains: Patents that reference scientific papers were drawn on especially heavily by future patents, and papers cited directly by patented inventions were especially highly cited by other scientific papers. Meanwhile, patents or papers that were disconnected from the other knowledge network were especially unlikely to be high impact within their own domains.

The impact advantages are robust to numerous controls, including fixed effects for each year, field, number of authors (paper) or inventors (patent), institution type, and each number of references made by the paper or patent (fig. S4). Fixed effect regressions account in a flexible and nonparametric manner for these features (see methods in supplementary materials). Tables S3 and S4 present the underlying regression results and also show that these results are robust to alternative measures of citation impact. We also find similar results using patent maintenance fee payments rather than citations received (table S5). Maintenance fees, which are paid by the patent owner and prevent the patent from lapsing, provide a potentially more direct measure of market value (30, 31). Figure S5 further shows that $D = 1$ patents did not simply cite established, popular papers; rather, papers cited by a patent in the year the paper was published tended to become home runs within science over the ensuing years. We also find that $D = 1$ patents and papers were far more likely to be home runs when looking within the outputs of a given inventor or author (tables S6 and S7). Examining individual fields, Fig. 3B shows that $D = 1$ patents and papers were the most highly cited within their own domains for the majority of scientific areas and technology classes. In science, 99% of fields, and in patenting, 86% of fields showed that the highest-impact work within the field occurs at $D = 1$.

Finally, we investigate the roles of institutions and individuals near the patent-paper boundary. Figure 4A considers institutions. For comparison,

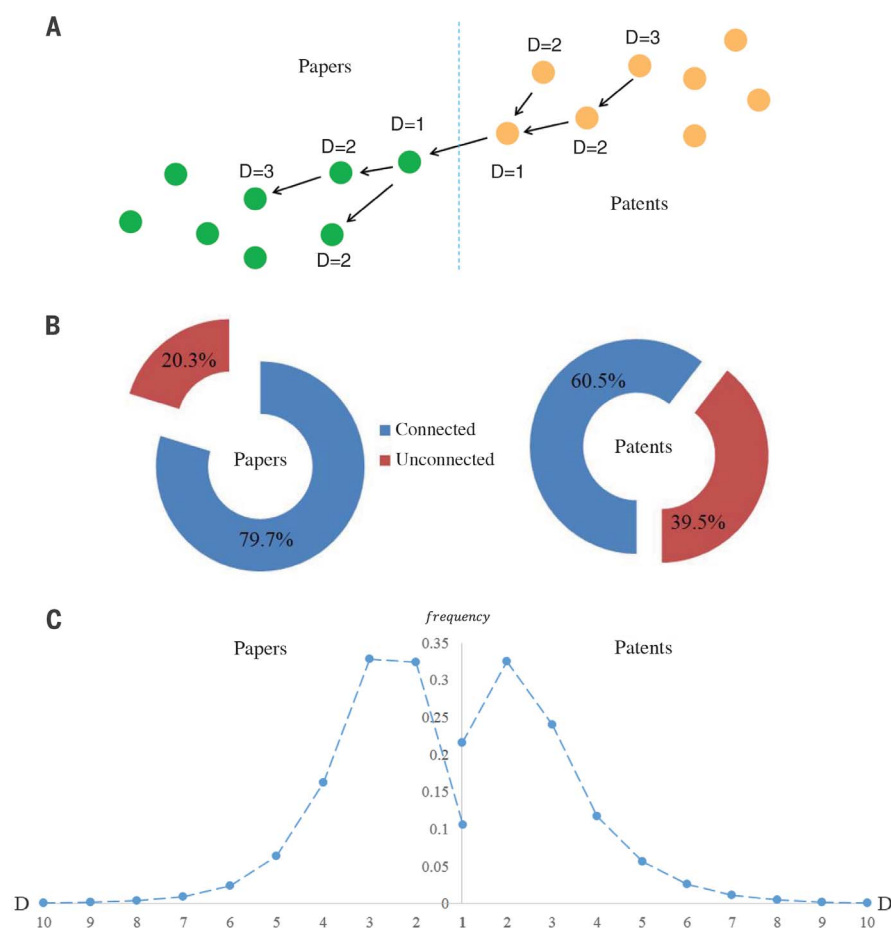


Fig. 1. Connectivity and distance. (A) The directed graph of the integrated citation network from patents toward papers defines a distance metric, D . (B) The share of papers that link forward to a future patent and the share of patents that link backward to a prior research article. (C) The distance distribution of connectivity.

we sorted relevant USPTO patents and WOS papers into three different institutional settings: universities, U.S. government laboratories, and firms. Institutional affiliations are based on patent assignee for the patents and based on postal and email addresses of the journal article authors (32, 33). The supplementary materials provide additional details of this sorting process. Universities and government laboratories were relatively more engaged in high- D research, whereas the research articles produced in firms shift toward $D = 1$ (Fig. 4A). These findings are consistent with and can help quantify long-standing ideas about the research outputs that for-profit institutions are likely to undertake (34). Table S8 provides associated regression analysis, including fixed effects for the number of references made, citations received, field, year, and number of authors or inventors. The regressions show that university papers were on average $D = 0.358$ further from the frontier than the firm papers. Decomposing this increased distance was due to field composition

(e.g., university researchers publish more in high- D fields such as mathematics than corporate researchers do) and two-thirds appeared as institutional differences within a given field (e.g., university papers in mathematics have higher D than firms' papers in mathematics).

Fully 57% of university-assigned patents had $D = 1$, indicating the intensiveness of university patenting near the boundary (Fig. 4A). Patents from firms peaked at $D = 2$, with only 19% at $D = 1$. Patents by government laboratories appeared in between the other institutions. Table S9 provides associated regression analysis, showing that, compared to firms, approximately one-half of university patents' increased proximity to science was due to field composition (university researchers patented in low- D technology classes) and one-half appeared as institutional differences within a given field (e.g., university patents in materials science had lower D than firms' patents in materials science).

We next considered the institutional "hand-off" across the boundary where $D = 1$. For $D = 1$ patents, 78% were assigned to firms, yet 80% of $D = 1$ papers had university authors (Fig. 4B). The prevalence of hand-offs from university papers to business patents is consistent with

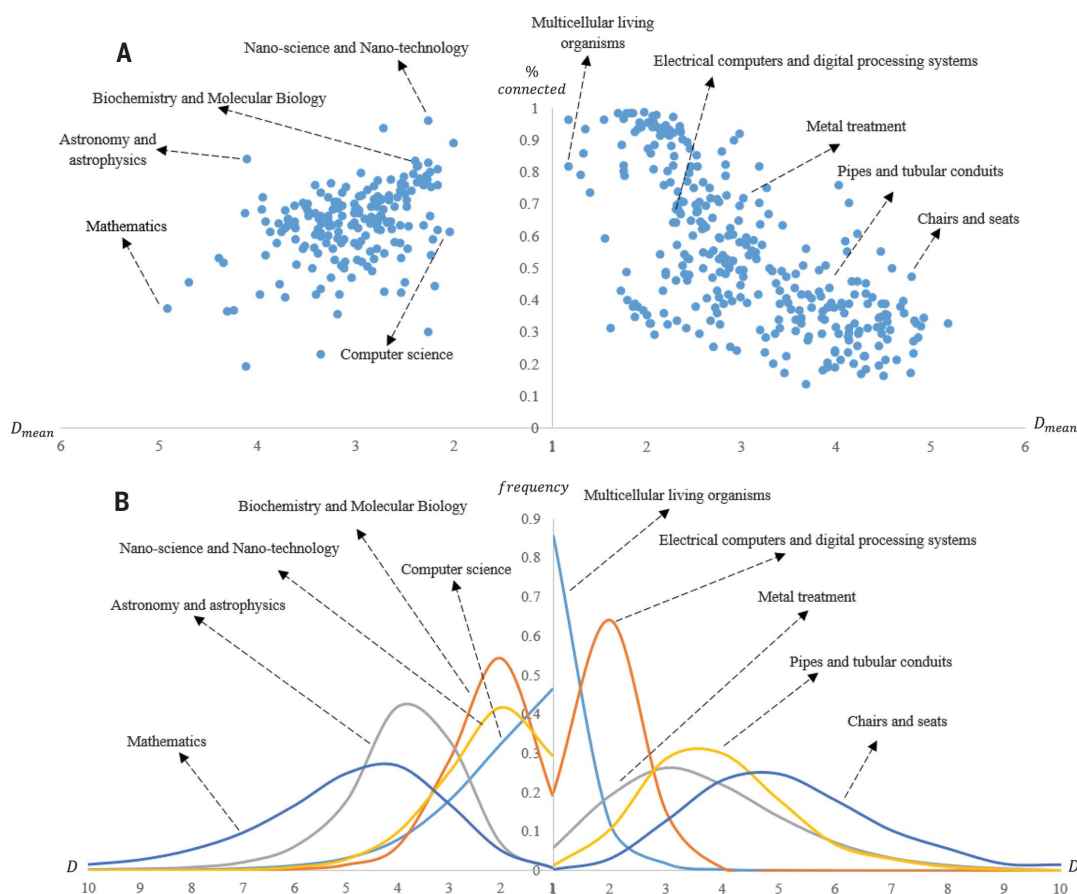


Fig. 2. Application to fields. (A) Distance metric. The mean distance, D_{mean} , to the paper-patent boundary is presented for each field (x axis) together with the percentage of knowledge outputs in that field that are connected to the integrated citation network (y axis). (B) The full D distribution for several fields.

long-standing conceptions that consider university outputs as public goods upon which marketplace invention can draw (1). Thus, although university patenting is particularly closely related to science (Fig. 4A) and can thus play a direct role in technology transfer (35, 36), the lion's share of $D = 1$ patents still comes from firms. Related, other patents typically connected to the patent-paper frontier through these $D = 1$ firm patents (fig. S6).

Figure 4C examines the role of the same individual in spanning the paper-patent boundary. We define these cases by matching the inventor names for the patent with the author names for the paper that the patent cites (see supplementary materials for further discussion). For $D = 1$ university patents, 55.4% cited a paper written by an individual with the same name. A high percentage also appeared for government patents, but the percentage fell to 14.3% for $D = 1$ corporate patents. In Stokes's theoretical characterization of "Pasteur's quadrant" (2), where the same individual may be engaged in advancing both understanding and use, universities and government laboratories appear to be especially common homes for such individuals, who in turn appear highly productive. Figure S7 and table S10 show that both the paper and the patent produced by such an individual

were especially likely to be home runs in their respective domains.

Contrary to conceptions in which technological and scientific progress operate in independent spheres, we find majority connectivity between the corpus of patented inventions and the corpus of scientific papers. However, these connections are typically indirect, and both scientific fields and patenting technology classes vary enormously in their connectivity and proximity to the other domain. These findings are consistent with and can help quantify some features of the "linear model" of science, which imagines that scientists typically work to advance understanding but that such advances may underlie practical applications, often in indirect or unexpected ways. The prevalence of private-sector patents linking back to the output of universities and government laboratories is further consistent with institutional views of the linear model. Although these features of the linear model appear to receive strong support, our data do not address potentially "nonlinear" reverse linkages where technological advances, including new equipment and tools, may also drive scientific progress (7, 11, 17).

The distance metric further reveals facts that are consistent with and help quantify the fruitful, creative interplay between understanding and

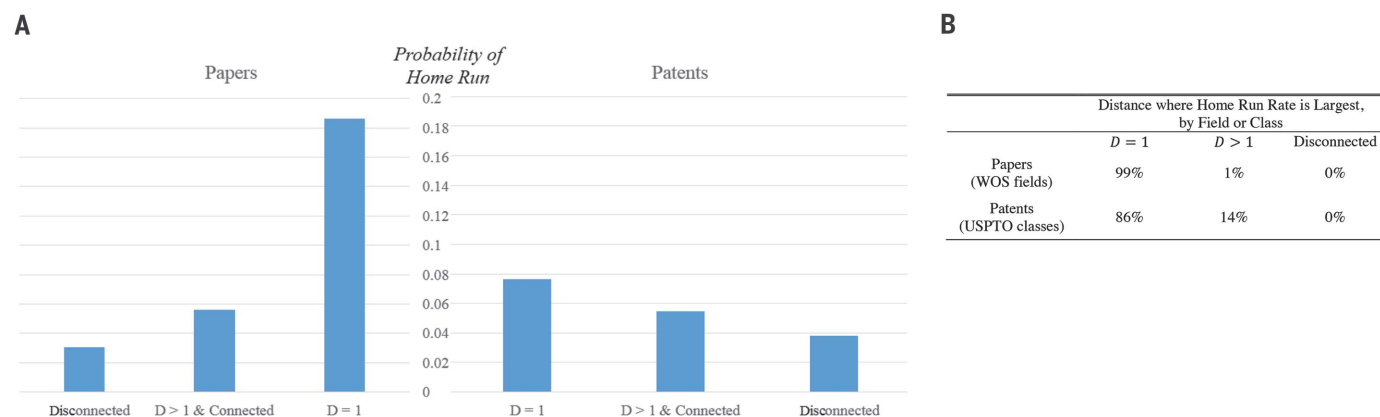


Fig. 3. Distance and impact. (A) Impact close to and far from the paper-patent boundary. A “home run” is defined as being in the upper 5% of citations received in that field and year, for a patent or a research paper.

(B) Home-run outcomes relative to distance for each field, when each field is analyzed separately. The supplementary materials examine alternative impact measures, including methods based on patent-renewal payments.

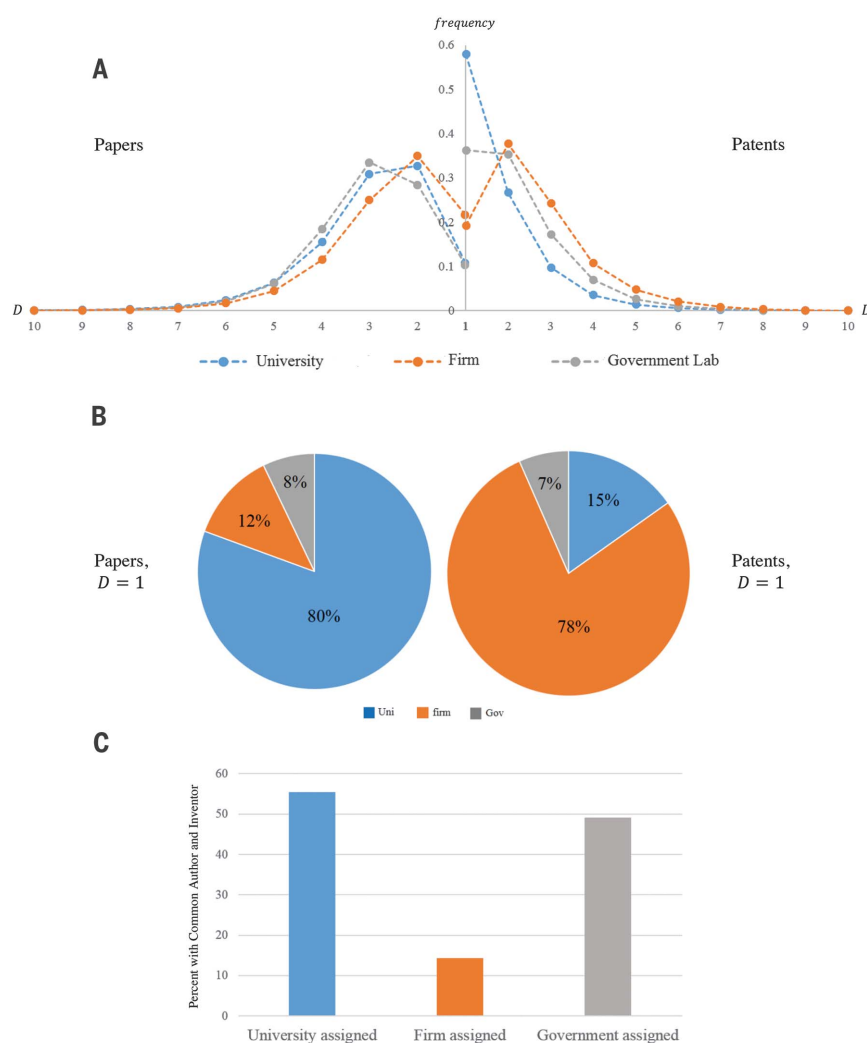


Fig. 4. Institutions and individuals. (A) The D distribution for different institutional settings, including universities, government laboratories, and firms. (B) Production of patents and papers by institutional type at the $D = 1$ boundary. (C) The share of $D = 1$ patents where a citing inventor and cited author have the same name, by patent assignee type.

application (2, 19, 21). Patented inventions that draw directly on scientific advances were especially impactful compared to other patents. Moreover, papers directly cited by patents were also the highest-impact papers within the scientific domain. These facts are consistent with a sharp complementarity between understanding and use and are also reflected at the individual level; an individual scientist/inventor, especially in university and government laboratory settings, often personally spanned the boundary, working to advance both the scientific and technological frontiers and managing to hit “home runs” in both domains.

Beyond loose classifications of “basic” or “applied” research and related terminologies (6, 7), the distance metric provides a quantifiable typology to describe R&D outputs and the nature of their impacts. The typology can characterize the research outputs of not only fields but also journals, funders, research institutions, and individuals themselves. Indices based on the D metric may thus present useful tools for understanding and evaluating types of research, institutional priorities, funding outcomes, and individual careers. While the distance metric in our application uses a directed graph, from patented invention to scientific advance, one may also deploy the metric on knowledge networks built using other link definitions. For example, full text analyses might allow one to characterize “necessary” precursor knowledge as opposed to the standard of “relevant” precursor knowledge that appears to be indicated by citation networks (see supplementary materials discussion). One might also build a metric that runs from scientific advances back to prior patented technologies, given appropriate reference information. And one might consider inventions or other applications outside patents. Such studies would further enrich our understanding of the interplay between scientific advance and technological progress to engage additional theories (11, 17).

REFERENCES AND NOTES

1. V. Bush, Science the Endless Frontier, A Report to the President (U.S. Government Printing Office, Washington, DC, July 1945).
2. D. Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation* (Brookings Press, Washington, DC, 1997).
3. J. Lane, S. Bertuzzi, *Science* **331**, 678–680 (2011).
4. J. Carty, *The Relation of Pure Science to Industrial Research*. Reprint and Circular Series, no. 14, National Research Council (1916).
5. W. R. Maclaurin, *Q. J. Econ.* **67**, 97–111 (1953).
6. B. Godin, *Sci. Technol. Human Values* **31**, 639–667 (2006).
7. M. S. Balconi, S. Brusoni, L. Orsenigo, *Res. Policy* **39**, 1–13 (2010).
8. National Science Foundation, *Basic Research: A National Resource* (U.S. Government Printing Office, Washington, DC, 1957).
9. R. Nelson, *J. Polit. Econ.* **67**, 297–306 (1959).
10. U.S. Department of Defense, Project Hindsight Final Report. Office of the Director of Defense Research and Engineering, Washington, DC (1969).
11. S. J. Kline, N. Rosenberg, in *The Positive Sum Strategy*, R. Landau, N. Rosenberg, Eds. (National Academy Press, Washington, DC, 1986), pp. 275–306.
12. E. Mansfield, *Res. Policy* **20**, 1–12 (1991).
13. A. Klevorick, R. Levin, R. Nelson, S. Winter, *Res. Policy* **24**, 185–205 (1995).
14. E. Von Hippel, *The Sources of Innovation* (Oxford Univ. Press, Oxford, 1988).
15. R. Nelson, in *The Rate and Direction of Inventive Activity: Economic and Social Factors*. National Bureau of Economic Research (Princeton Univ. Press, Princeton, NJ, 1962), pp. 549–583.
16. N. Rosenberg, in *Inside the Black Box: Technology and Economics* (Cambridge Univ. Press, Cambridge, 1982), pp. 141–159.
17. H. Brooks, *Res. Policy* **23**, 477–486 (1994).
18. F. Narin, K. Hamilton, D. Olivastro, *Res. Policy* **26**, 317–330 (1997).
19. L. Fleming, O. Sorenson, *Strateg. Manage. J.* **25**, 909–928 (2004).
20. R. Gaetani, M. Li Bergolis, The economic effects of scientific shocks. Northwestern University working paper (2015).
21. B. Cassiman, R. Veugelers, P. Zuniga, *Ind. Corp. Change* **17**, 611–646 (2008).
22. A. B. Jaffe, B. Trajtenberg, *Patents, Citations, and Innovations: A Window on the Knowledge Economy* (MIT Press, 2002).
23. M. Meyer, *Res. Policy* **29**, 409–434 (2000).
24. R. J. W. Tijssen, R. K. Buter, Th. N. Van Leeuwen, *Scientometrics* **47**, 389 (2000).
25. D. Hicks, T. Breitzman, D. Olivastro, K. Hamilton, *Res. Policy* **30**, 681–703 (2001).
26. J. Callaert, M. Pellens, B. Van Looy, *Scientometrics* **98**, 1617–1629 (2014).
27. S. Wuchty, B. F. Jones, B. Uzzi, *Science* **316**, 1036–1039 (2007).
28. B. F. Jones, S. Wuchty, B. Uzzi, *Science* **322**, 1259–1262 (2008).
29. D. Wang, C. Song, A. L. Barabási, *Science* **342**, 127–132 (2013).
30. M. Schankerman, A. Pakes, *Econ. J. (Oxf.)* **96**, 1052–1076 (1986).
31. D. Harhoff, F. Narin, F. M. Scherer, K. Vopel, *Rev. Econ. Stat.* **81**, 511–515 (1999).
32. A. Arora, S. Belenzon, Killing the golden goose? The decline of science in corporate R&D. NBER Working Paper No. 20902 (2015).
33. K. Bryan, Y. Ozcan, The impact of open access mandates on invention. University of Toronto Working Paper (2016).
34. N. Rosenberg, R. Nelson, *Res. Policy* **23**, 323–348 (1994).
35. National Academy of Sciences, *Managing University Intellectual Property in the Public Interest* (National Academy Press, Washington, DC, 2010).
36. J. G. Thursby, M. C. Thursby, *Science* **301**, 1052 (2003).

ACKNOWLEDGMENTS

We gratefully acknowledge support from the Alfred P. Sloan Foundation under award G-2015-14014 and support from the Northwestern Institute on Complex Systems. We thank workshop participants at the National Academy of Sciences, the European Policy for Intellectual Property Association, the Northwestern Institute for Complex Research, the Institute for Policy Research, P. Azoulay, A. Jaffe, A. Marco, M. Trajtenberg, and B. Uzzi for helpful comments. We are especially grateful to R. Gaetani for help with patent data. The Web of Science data are available via Thomson Reuters. The patent data sets are available publicly as discussed in the supplementary materials. The constructed *D* metric variables are available from the authors: bjonas@kellogg.northwestern.edu (corresponding author) and mohammad.ahmadpoor@kellogg.northwestern.edu.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/357/6351/583/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S10
References (37–52)

9 February 2017; accepted 10 July 2017
10.1126/science.aam9527

The dual frontier: Patented inventions and prior scientific advance

Mohammad Ahmadpoor and Benjamin F. Jones

Science **357** (6351), 583-587.
DOI: 10.1126/science.aam9527

Picking up a patent

What is the relationship between patents and scientific advances? Ahmadpoor and Jones devised a metric for the "distance" between patentable inventions and prior research to study this question. They analyzed the relationship between 4.8 million U.S. patents and 32 million research articles. Universities tended to cite their own research directly in their patents (in other words, a distance of 1), but the distance was greater for companies, suggesting that companies may rely on outsiders for their foundational research. The distance varied by discipline, with nanotechnology and computer science having the shortest distances between published research and patents.

Science, this issue p. 583

ARTICLE TOOLS

<http://science.sciencemag.org/content/357/6351/583>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2017/08/09/357.6351.583.DC1>

REFERENCES

This article cites 35 articles, 7 of which you can access for free
<http://science.sciencemag.org/content/357/6351/583#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)