Harnessing Academic Science for Corporate Technology: Interpersonal Networks and Absorptive Capacity[1]

Sam Arts
Department of Management, Strategy and Innovation
Faculty of Economics and Business
KU Leuven
sam.arts@kuleuven.be

Lee Fleming
Management of Organizations, Haas School of Business
Industrial Engineering and Operations Research, College of Engineering
UC Berkeley
lfleming@berkeley.edu

Elena Veretennik
Department of Management, Strategy and Innovation
Faculty of Economics and Business
KU Leuven
elena.veretennik@kuleuven.be

ABSTRACT

If firms conduct less scientific research internally and yet increasingly rely on science for innovation, how do they gain access to external scientific knowledge? To examine the role of interpersonal networks among corporate inventors and academic scientists in facilitating scientific knowledge transfer from academia to industry, we construct a social network encompassing nearly the entire population of life science academic scientists and corporate inventors listed on U.S. patents. To isolate the influence of interpersonal connections from the inherent characteristics and commercial potential of scientific discoveries, we leverage paper twins—scientific papers with nearly identical findings published around the same time by independent academic teams—and analyze their citations in corporate patents. Although academic science is often viewed as a public good, our findings highlight the role of close interpersonal connections—within two degrees of separation—in harnessing academic science for corporate innovation. Moreover, corporate inventors' ability to leverage these connections depends critically on their own active engagement in scientific research as well as the alignment of their scientific expertise with the academic knowledge they seek to integrate.

[1] Acknowledgements: We gratefully acknowledge comments from seminar participants at Copenhagen Business School, KU Leuven, the Max Planck Institute for Innovation and Competition, Pompeu Fabra University, Rotterdam School of Management, Scuola Superiore Sant'Anna, and Warwick Business School, as well as attendees of the 9th Summer School on Data and Algorithms for ST&I Studies, AOM 2025, DRUID 2024, the 7th RISE Workshop in Munich, and WOEPS 2025. We are especially thankful to Paola Criscuolo, Fabian Gaessler, Daehyun Kim, Amit Kumar, Nicola Melluso, Ali Mohammadi, Felix Poege, Chandrika Rathee, Cesare Righi, Ammon Salter, Sotaro Shibayama, and Markus Simeth for their comments and suggestions.

1. Introduction

Scientific research has long been recognized as a key driver of technological progress in industry (Bush 1945, Mokyr 2002). The "scientification" of corporate technology is particularly evident in industries such as biotechnology and semiconductors, where firms increasingly rely on scientific research for innovation (Henrekson and Rosenberg 2001). Despite this growing reliance, firms—particularly large ones—have scaled back internal investments in scientific research, opting instead to increasingly rely on external scientific knowledge from academia (Arora et al. 2018, Fleming et al. 2019). Although extensive research demonstrates that relying on academic science enhances corporate innovation and financial performance, the specific mechanisms that enable firms to identify, access, and integrate external scientific knowledge remain less well understood (Cockburn and Henderson 1998, Zucker et al. 2002, Fleming and Sorenson 2004). While prior research has extensively examined the transfer of academic knowledge to industry through formal market mechanisms such as licensing, contracting, and the creation of academic startups (Jensen and Thursby 2001, Shane 2004, Agrawal 2006), other work suggests that such knowledge transfers are inherently social processes, occurring primarily through interpersonal connections rather than being exclusively governed by market transactions (Liebeskind et al. 1996, Murray 2004, Perkmann et al. 2013, Perkmann et al. 2021).

Accessing and leveraging external academic science poses several challenges. The rapid expansion and dispersion of scientific knowledge—millions of new papers annually—makes it difficult for firms to search and filter relevant scientific advancements. Moreover, much of this scientific knowledge is complex and partly tacit—residing in part in the minds of the academic scientists who produce it—complicating its transfer through publications alone (Collins 1974, Zucker et al. 2002, Roche 2023). The ability of corporate inventors to access and integrate these scientific insights may therefore depend on their interpersonal connections with the academic scientists responsible for the scientific discoveries (Murray 2004, Agrawal 2006). Connections alone may not suffice, however. Corporate inventors might also require personal absorptive capacity—the scientific expertise needed to recognize, assimilate, and apply external academic knowledge in the development of corporate technologies (Cohen and Levinthal 1990, Pavitt 1991, Cockburn and Henderson 1998, Cavallo et al. 2025).

This paper explores how firms integrate academic science by focusing on the interaction between interpersonal networks linking corporate inventors to academic scientists and inventors' absorptive capacity. We use corporate patent citations to academic papers as an indicator of knowledge integration (Fleming and Sorenson 2004, Arora et al. 2018) and construct a large-scale collaboration network linking over 9 million PubMed authors—covering the near-entire population of academics in life sciences—and 2.5 million U.S. patent inventors. This new dataset includes a unique crosswalk that, for the first time, enables the measurement of social proximity between any academic scientist in life sciences and any

corporate inventor (Torvik 2018). While prior work has focused on direct collaborations—typically captured through co-authorships between academic scientists and corporate inventors (e.g., Liebeskind et al. 1996, Cockburn and Henderson 1998)—we leverage this new dataset to adopt a broader, network-based perspective. We examine the full structure of interpersonal networks, including indirect connections, that may facilitate firms' access to and integration of scientific knowledge. We also examine how social proximity and inventors' absorptive capacity interact in shaping the integration of academic science. To do so, we test whether corporate inventors are better able to leverage social proximity to academic scientists when they remain actively engaged in scientific research and when their prior scientific expertise aligns with the academic knowledge they seek to integrate.

To isolate the influence of network connections from the intrinsic characteristics and commercial potential of scientific discoveries, we employ academic paper twins—nearly identical scientific findings published around the same time by independent academic teams (Merton 1961, Bikard et al. 2019). We then identify corporate patents that cite only a subset of the twin papers, using the uncited twin as a counterfactual to estimate the effect of interpersonal ties on knowledge transfer (Bikard 2018). Applying conditional logit models, we estimate the effect of social proximity—measured as the shortest path between corporate inventors and academic scientists—on the probability of citation, while controlling for unobserved heterogeneity in the scientific discovery, patented technology, and the patenting firm and team of corporate inventors.

While academic science is generally viewed as a public good—freely accessible and beneficial to all without exclusion or competition (Nelson 1959)—our findings suggest that firms are better able to tap into external scientific knowledge when their inventors are socially connected to the academic scientists who produced that science. The probability of knowledge transfer declines significantly and monotonically as social distance within the network increases. Indirect connections can still facilitate knowledge transfer, though their effect becomes negligible beyond two degrees of separation—that is, when a corporate inventor is linked to the academic scientist who made the scientific discovery only via a common collaborator. Importantly, however, interpersonal ties do not appear to operate independently of inventors' absorptive capacity: social proximity matters primarily when corporate inventors remain actively engaged in scientific research, and its value is further amplified when their scientific expertise aligns with the academic knowledge they seek to apply (Rosenberg 1990, Cohen and Levinthal 1990). Finally, replicating our design using paper-to-paper instead of patent-to-paper citations to the same twin papers, we find that interpersonal proximity also shapes science-to-science knowledge transfer, but significantly less so. This contrast suggests that interpersonal proximity plays a larger role when scientific knowledge is translated into corporate technology than when it circulates within academia.

Our findings hold across both large public firms and smaller firms, including startups, and remain robust to a range of checks, including alternative measures of social distance. To address concerns that our results may be driven by endogenous tie formation for knowledge transfer, we demonstrate robustness by excluding observations in which academic scientists and corporate inventors have direct connections, as these cases are most likely to reflect intentional tie formation. In addition, our results remain consistent when social distance is measured solely based on co-authorship links established prior to the publication of the academic twin papers, thereby eliminating ties that may have (intentionally) formed after the discovery specifically to facilitate knowledge transfer and commercialization.

Our study makes two main contributions. First, we ground our work in the micro-foundations literature, which emphasizes that firm-level constructs such as network advantages and absorptive capacity ultimately arise from the actions and interactions of individuals (Felin et al. 2012). Building on this perspective, we demonstrate how individual inventors—through their ties to academic scientists and their own absorptive capacity—shape their firms' capability to access and integrate external scientific knowledge. In doing so, we advance research on the value of academic science for firms (e.g., Fleming and Sorenson 2004, Simeth and Cincera 2016) by shifting attention from the outcomes of using science to the mechanisms that enable firms to source and absorb it (Bikard et al. 2019). To our knowledge, this is the first empirical evidence showing that interpersonal networks—including both direct and indirect ties up to two degrees—between academic scientists and corporate inventors play a key role in firms' ability to access and integrate academic knowledge. Whereas prior research has focused on direct interpersonal collaboration (e.g., Cockburn and Henderson 1998, Zucker et al. 2002), we adopt a network-based perspective and show that a greater number of knowledge transfers from academia to industry occur through indirect connections, consistent with network theory on the value of indirect ties in providing access to novel, non-redundant, and distributed information (Granovetter 1973, Cowan et al. 2007). Moreover, while prior work shows that interpersonal networks shape the diffusion of scientific knowledge among academic scientists (Mohnen 2022) and technological knowledge among corporate inventors (Singh 2005, Paruchuri and Awate 2017, Moreira et al. 2018), we leverage an integrated network that links academic scientists and corporate inventors. This mixed network of scientists and inventors allows us to show how interpersonal connections across the academia–industry boundary help facilitate the transfer of scientific knowledge into corporate technology.

Second, we contribute to classic work on boundary spanning (Allen 1977, Tushman 1977, Tushman and Scanlan 1981) and individual absorptive capacity (Cohen and Levinthal 1990) by theorizing and demonstrating that corporate inventors benefit from interpersonal ties to academia primarily when they remain actively engaged in both scientific research and technology development. These dual-engaged inventors are better positioned to recognize, assimilate, and apply external scientific insights from socially

connected academics—especially when their scientific expertise closely aligns with the academic work they seek to integrate (Scharfmann et al. 2025). While prior work has shown that corporate inventors involved in scientific research contribute to firm innovation (Cockburn and Henderson 1998, Zucker et al. 2002, Gittelman and Kogut 2003, Arts and Veugelers 2020, Schaper et al. 2025), our study identifies one mechanism underlying this advantage: interpersonal ties that provide social access to academic knowledge, combined with inventors' absorptive capacity that enables them to recognize, interpret, and apply that knowledge in corporate technology.

2. Theory and Hypotheses

Firms increasingly rely on academic science to fuel technological innovation, particularly as they scale back internal investments in scientific research (Arora et al. 2018, Fleming et al. 2019). Yet despite the growing importance of external scientific knowledge, integrating academic science into corporate technology remains challenging. Prior work has mostly examined formal market mechanisms for transferring academic knowledge to industry, such as licensing, sponsored research contracts, and university spinoffs (Jensen and Thursby 2001, Shane 2004). However, a growing body of work suggests that knowledge exchange between academia and industry is fundamentally a social process, facilitated by interpersonal connections and interactions rather than mediated exclusively through market transactions (Liebeskind et al. 1996, Owen-Smith and Powell 2004). While ongoing collaborations between academic scientists and corporate inventors likely provide the strongest conduit for knowledge transfer, prior collaborative ties often remain effective channels for continued knowledge flow (Cowan et al. 2007, Fleming et al. 2007).

Interpersonal networks between academic scientists and corporate inventors can foster knowledge exchange across the academia–industry boundary in several ways. First, interpersonal networks can help corporate inventors become aware of relevant scientific insights (Cockburn and Henderson 1998, Murray 2004). Given the overwhelming volume of external academic research, such network ties may reduce search costs by filtering and directing attention to relevant and high-quality scientific findings. Second, interpersonal ties can improve access to contextual and partly tacit knowledge, including know-how, techniques, and insights that are not fully codified in publications but may be important for developing technological applications (Collins 1974, Zucker et al. 2002, Agrawal 2006). An inventor connected to the academic who made the scientific discovery, or to their collaborators, may obtain informal insights or practical guidance not captured in the published work. Third, interpersonal contact can build trust and credibility, making corporate inventors more confident in the validity and importance of academic findings and increasing academics' willingness to share novel insights that might otherwise remain difficult to access across institutional boundaries (Owen-Smith and Powell 2004).

Both direct and indirect ties are expected to contribute to these processes. Direct ties may foster mutual awareness, trust, and credibility, and provide richer opportunities for exchanging contextual and partly tacit information (Collins 1974, Hansen 1999, Zucker et al. 2002). Indirect ties—such as those formed through mutual collaborators—can also facilitate knowledge diffusion and may provide access to more diverse and novel information than direct connections (Granovetter 1973, Cowan et al. 2007). A corporate inventor may not have collaborated personally with a given academic but may be linked through a shared collaborator. Such second- and higher-degree connections may still enable informal exchanges of research insights through various interaction channels and professional networks where mutual contacts bridge otherwise distant actors. Crucially, such knowledge flows do not necessarily require direct interaction: a mutual collaborator may flag a new scientific finding, circulate a paper, or contextualize its relevance for an ongoing technological problem. Even brief encounters or indirect referrals of this kind can raise awareness of relevant scientific discoveries and offer contextual information about their potential commercial application. Such indirect channels may, in turn, serve as a catalyst for deeper engagement—for instance, prompting subsequent informal advice exchanges or targeted problem-solving discussions—through which scientific insights may ultimately be assessed and integrated into technological development. Moreover, while each indirect tie is individually less likely to facilitate knowledge transfer than a direct tie, there are often many more indirect than direct ties, and in aggregate they may transmit a greater volume of information.

Despite their aggregate importance, the effectiveness of indirect ties diminishes as network distance increases (Singh 2005). Distant ties may suffer from lower interaction frequency, reduced trust, or fewer opportunities to communicate substantive knowledge. Thus, we expect closer ties between academic scientists and corporate inventors to be more strongly associated with the incorporation of academic science into corporate technology. Indirect ties should also increase the likelihood of building upon academic science, but this influence should weaken as the number of intermediaries—that is, the social distance between inventors and academic scientists—increases.

Hypothesis 1: Corporate technology is more likely to build on academic science as corporate inventors are more closely connected to the academic scientists who produced that science.

Social proximity alone may not guarantee the successful application of academic science to corporate invention, because mere exposure to external scientific knowledge may not be sufficient for inventors to integrate it effectively. Corporate inventors likely vary in their ability to recognize, assimilate, and apply academic science. We therefore propose two interaction hypotheses, suggesting that inventors' absorptive capacity moderates the effect of social proximity on the use of academic science. Specifically, and consistent with absorptive capacity theory, we distinguish two dimensions of inventors' absorptive

capacity: active engagement in scientific research at the time of corporate invention, and alignment between an inventor's scientific expertise and the focal academic research.

Active engagement in scientific research enables corporate inventors to develop and maintain personal absorptive capacity—that is, the ability to recognize the value of new external scientific knowledge, assimilate it, and apply it in the development of commercial technology (Cohen and Levinthal 1990). Such inventors are more likely to notice new academic findings and understand their implications for the firm's innovation projects (Anckaert and Cassiman 2026). In contrast, inventors without active scientific engagement may not immediately grasp an academic discovery, even if an academic colleague draws attention to it. Even with interpersonal connections, they may lack the absorptive capacity needed to make sense of what they learn through those connections (Tushman and Scanlan 1981).

If these arguments are correct, corporate inventors' active involvement in scientific research should amplify the benefits of network ties, above and beyond the value of network connections alone (Operti and Carnabuci 2014). Corporate inventors who actively conduct scientific research are more effectively bilingual in the languages of science and technology (Cockburn and Henderson 1998). They can read academic papers and translate new scientific insights for corporate colleagues who are not actively engaged in scientific research. In essence, corporate inventors who straddle both worlds act as boundary spanners or gatekeepers, translating and importing external knowledge from academia into the firm (Allen 1977, Tushman 1977).

Moreover, active scientific engagement embeds corporate inventors in the academic community to some degree, increasing their visibility and credibility (Rosenberg 1990). This can make academics more willing to share information or collaborate, further strengthening knowledge exchange. At the organizational level, prior work shows that firms with such boundary-spanning scientist-inventors generate more innovation (Cockburn and Henderson 1998, Zucker et al. 2002, Arts and Veugelers 2020). We posit that one reason is their ability to exploit social ties to academia more effectively. For these reasons, the positive effect of social proximity to academic scientists should be stronger for corporate inventors who actively engage in scientific research.

Hypothesis 2: The increased likelihood that corporate inventors with social proximity will build on academic science will be greater when inventors are actively engaged in scientific research.

Independent of the benefits of active engagement in scientific research, corporate inventors' ability to leverage social networks may also depend on how closely their scientific expertise aligns with the academic knowledge they seek to integrate. Scientific knowledge is often highly specialized, complex, and domain-specific (Collins 1974), making it difficult to identify, absorb, and apply in commercial technology development without the appropriate background and expertise (Ellis 1965). Prior work

suggests that individuals' absorptive capacity is stronger when it is domain-specific: individuals are better able to recognize, assimilate, and apply new external knowledge when that knowledge relates closely to their prior experience and expertise (Cohen and Levinthal 1990, Lane and Lubatkin 1998).

Corporate inventors with closely related scientific expertise are more likely to recognize valuable academic insights, understand their technological implications, and translate them into commercial applications. Crucially, the effectiveness of social proximity to academic scientists depends on this alignment. When corporate inventors possess scientific expertise closely related to the focal academic domain, their interpersonal ties to academic scientists are more likely to support meaningful knowledge exchange. Conversely, when corporate inventors lack relevant scientific expertise, even close interpersonal connections may be insufficient for effective knowledge integration.

Hypothesis 3: The increased likelihood that corporate inventors with social proximity will build on academic science will be greater when inventors' scientific expertise is closely related to the focal academic science.

Together, these arguments imply two related but distinct interaction effects that reflect key dimensions of individual absorptive capacity: active engagement in scientific research and closely related scientific expertise each strengthen the relationship between social proximity and the integration of academic science into corporate technology.

3. Data and Empirical Strategy

3.1 Collaboration network spanning PubMed authors and U.S. patent inventors

To construct the collaboration network linking all PubMed authors and U.S. patent inventors, we use disambiguated author and inventor data from Torvik and Smalheiser (2009, 2018), Smalheiser and Torvik (2009), and Torvik (2018), covering records through 2009. The PubMed author disambiguation dataset is widely used in prior research (e.g., Azoulay et al. 2017) and includes linked files identifying both PubMed authors and U.S. patent inventors, allowing us to assign unique person identifiers and reconstruct each individual's co-authored papers and patents.

First, we construct a bipartite graph consisting of instances (papers and patents) and individuals (their authors and inventors). Each co-authored paper or patent creates a link between the individuals involved from the year of publication or patent filing onward (e.g., Breschi and Catalini 2010). We treat the publication or filing year as the establishment date of a co-authorship tie (e.g., Singh 2005). We then project this bipartite graph into a unimodal network in which nodes represent individuals and edges indicate prior co-authorship at a given point in time. Each new co-authored paper or patent may thus create collaboration links between previously unconnected individuals.

To capture the dynamic nature of the network, we construct a separate network for each year from 1975 to 2009, based on all co-authorship ties formed up to and including that year. The final 2009 network includes 11.6 million individuals—8.99 million authors, 2.34 million inventors, and 0.3 million inventor-authors—and 80.2 million direct collaboration links, constructed from 18.6 million papers and 3.9 million patents. Using this network, we identify interpersonal ties and social proximity between any corporate inventor and academic scientist at a given point in time. Following prior research, we measure social proximity using shortest path length, defined as the minimum number of edges required for knowledge to flow between two individuals (Wasserman and Faust 1994). This measure captures whether two individuals are socially connected and, if so, how many steps separate them. In robustness checks, we show that our results hold when using an alternative measure of social distance that accounts for tie decay over time.

3.2 Academic paper twins to strengthen identification

To separate the influence of interpersonal networks on bridging academic science and corporate technology from the influence of the content and commercial potential of the scientific discoveries themselves, we leverage instances of simultaneous academic discoveries. This identification approach involves identifying cases of "paper twins" where nearly identical findings are published around the same time by independent academic teams as described by Merton in 1961 and previously used in empirical research as an identification strategy (Bikard 2018, Bikard et al. 2019, Bikard and Marx 2020, Marx and Hsu 2022). We use an open-access dataset of paper twins from Bikard (2020), which identifies pairs of PubMed papers that 1) are in the top 10 of most similar papers according to PubMed's most similar article list, 2) do not have any common authors, 3) are not published more than one year apart, (4) received at least five citations each and the Jaccard index based on the overlap in the citing papers is at least 50%, i.e., the papers are mostly cited together by the same papers, 5) and finally at least three co-citing articles must be available in open access for credit allocation check, i.e., to check whether they are cited adjacently or within the same parentheses. The co-citation reflects that the scientific community gives equal credit to twin papers by citing them together (Merton 1957). Below, we illustrate the robustness of our findings in case we use stricter criteria to identify paper twins.

Starting from the full sample of potential paper twins from Bikard (2020), we retain only those in which all authors are affiliated with universities or public research institutions, excluding any paper co-authored by individuals with corporate affiliations.[2] Following the identification approach of prior studies, we keep the subset of paper twins where a company patent is only citing a subset of the twin papers

[2] Affiliation data was collected and verified across three sources: Web of Science, PubMed, and the "MapAffil" dataset by Torvik (2021).

(Bikard 2018, Bikard and Marx 2020, Marx and Hsu 2022).[3] These criteria resulted in a sample of 247 simultaneous discoveries comprising 501 academic papers published between 1975 and 2007. Some papers were paired with multiple others, forming seven triplets.[4] The twin papers in our dataset are most often affiliated with universities such as Harvard, MIT, UC San Diego, and UC San Francisco, and are mostly published in top journals like *Nature*, *PNAS*, *Cell*, and *Science* (see Appendix Table A.1). The twin papers in our sample are highly similar, with a median publication gap of 2.5 months (average: 2.49) and a median PubMed rank difference of 3.5 (average: 6.27). Using SPECTER embeddings, we find a median cosine similarity of 0.92 (average: 0.89), confirming almost identical scientific content (Cohan et al. 2020). The median Jaccard index of shared citations is 0.55 (average: 0.57), and 81% (average: 76%) of citing papers mention both twins adjacently. Notably, 90 of the 247 twin papers were published back-to-back in the same journal issue, a common signal of simultaneous discovery.

Figure A.1 in the Appendix shows an example of a paper twin from our data. In 1998, two independent teams of academic scientists simultaneously discovered that the CD40 signal can activate killer immune cells independently of helper cells (Bennett et al. 1998, Schoenberger et al. 1998). Published back-to-back in *Nature*, these twin papers are ranked first in PubMed's list of most similar papers, with a cosine similarity of 0.97, indicating virtually identical content. Each paper has received over a thousand citations, and a 77% adjacent co-citation rate, reflecting equal credit allocation. This simultaneous scientific discovery was a breakthrough that paved the way for developing new vaccines and treatments for those with weakened immune systems, such as AIDS patients, and has potential applications in cancer treatment. The substantial commercial value is reflected in over 250 patents that cite these scientific publications. Among them, approximately 20% cite only the Schoenberger paper, 15% cite only the Bennett paper, and the remaining patents cite both.

3.3 Variables and estimation method

We identify all patent citations to twin papers from U.S. utility patents filed by firms between 1975 and 2009 (Marx and Fuegi 2020). Following our identification strategy, we include only cases in which a patent cites one (or a subset, in the case of triplets) of the twin papers. This design allows us to use the uncited twin paper to construct a counterfactual or "pseudo-citation" and estimate whether social distance between inventors and authors influences citation likelihood using conditional logit models. Because twin papers report essentially the same scientific discovery, inventors could, in principle, have cited either or both. U.S. patent law requires disclosure of all relevant prior art, and non-disclosure risks rejection or

[3] However, as we demonstrate later, our results remain consistent even when using the full sample, which includes company patents citing all papers from the same twin discovery.
[4] Triplets refer to a unique grouping of three papers conducted by independent research teams.

invalidation of the patent. It is therefore unlikely that inventors were aware of the uncited twin paper but deliberately chose not to cite it. Supporting this interpretation, Bikard and Marx (2020) conducted 21 interviews with inventors citing paper twins and found that inventors themselves—rather than patent attorneys or examiners—were usually responsible for adding references to scientific literature, a finding consistent with an earlier survey (Roach and Cohen 2013). The inventors emphasized that citations are driven more by their familiarity with specific papers than by legal necessity or an intention to cite all relevant prior art. Nevertheless, our results remain robust when we extend the analysis to include cases in which corporate patents cite all papers from the same twin discovery.

Our unit of analysis is a paper–patent dyad, encompassing both cited paper-patent dyads (real patent-to-paper citations) and uncited paper-patent dyads (counterfactual or pseudo patent-to-paper citations). Our final sample includes 247 academic paper twins, 501 papers, and 1,951 citing company patents. This results in 4,169 paper-patent dyads, of which 2,070 cited paper-patent dyads and 2,099 are uncited paper-patent dyads. The academic twin papers are most frequently cited in patents from companies such as Genentech, Millennium Pharmaceuticals, Zymogenetics, and BASF Plant Science (see Table A.2 in the Appendix).

To study whether interpersonal networks bridge academic science and corporate technology, we follow the approach of Bikard (2018) and estimate the following conditional logit model with paper twin-patent level fixed effects:

$$cited_{ijk} = \alpha_0 + \alpha_1 * social_dist_{ik} + \alpha_2 * X_{ik} + \alpha_3 * Y_i + \gamma_{jk} + \varepsilon_{ijk}$$

In this model, $cited_{ijk}$ is a binary outcome variable equal to 1 in case paper *i* from paper twin *j* is cited by patent *k* and 0 in case patent *k* is not citing paper *i* but another paper of the same twin *j*. Our main explanatory variable is $social_dist_{ik}$ which captures the shortest path, i.e., the minimum number of edges in the collaboration network, between any of the academic authors of paper *i* and any of the company inventors of patent *k.* In a small percentage of the paper-patent dyads (1.7%), there is no connection between any of the inventors and authors and $social_dist_{ik}$ has a missing value. In these cases, we assign a value of 9 as 8 is the maximum observed value in our dataset. Nevertheless, we find consistent results dropping observations with missing values from the analysis. To avoid imposing a linear functional-form assumption, we also use a series of binary indicators to capture social distance. For instance, *social_dist_0* is a binary indicator equal to 1 if the shortest path is 0, i.e., an author of paper *i* is the same person as an inventor of patent *k*. As another example, *social_dist_1* is a binary indicator equal to 1 if the shortest path is 1, i.e., an author of paper *i* is a direct prior collaborator of an inventor of patent *k*.

To test the interaction effect outlined in *Hypothesis 2*, we first assess whether corporate inventors listed on a patent are actively engaged in scientific research at the time of patent filing. We define a binary variable, *active_inventor_author*, which equals one if any co-inventor of a patent has published a

scientific paper in the year preceding the patent's filing.[5] As discussed below, we control for the team's overall prior patenting and publishing to ensure that the *active_inventor_author* variable captures simultaneous engagement in scientific research during corporate technology development. To examine how corporate inventors' active engagement in scientific research enables them to leverage interpersonal connections with academic scientists, we introduce an interaction term between *social_dist* and *active_inventor_author* in our model.

To examine the interaction effect proposed in *Hypothesis 3*, we measure the degree to which corporate inventors' scientific expertise aligns with the academic research they build upon. To quantify this alignment, we compute SPECTER embeddings for each paper based on its title and abstract and calculate the cosine similarity between corporate inventors' prior publications and the academic twin papers. At the twin–patent level, we take the maximum of these pairwise similarity scores as a proxy for the extent to which the inventor team possesses scientific expertise closely related to the focal twin discovery. We then define a binary variable, *close_paper*, set to one if this similarity measure exceeds the median value in our dataset.[6] Finally, to evaluate whether closely related scientific expertise enhances corporate inventors' ability to leverage interpersonal connections with academic scientists, we incorporate an interaction term between *social_dist* and *close_paper* in our model.[7]

Turning to control variables, we note that although twin papers capture nearly identical scientific findings, they remain distinct publications authored by different teams. The characteristics at both the paper and team levels may affect the likelihood of a paper being cited by a corporate patent. To account for this, Y_i in the model above represents a vector of paper-level control variables that might influence the likelihood that paper *i* is cited by patent *k*. We include *no. authors* (number of authors on the paper), *degree3_size* (number of individuals within 3 degrees of separation of the paper's authors up to the year of patent filing, i.e., to control for the size of the authors' social network), *author-inventor team* (binary indicator equal to 1 if at least one of the paper's authors is an author-inventor at the patent filing year), *publication stock* (number of unique PubMed papers published by any of the paper's authors up to the

[5] We find consistent results when defining *active_inventor_author* as equal to one if any co-inventor of a patent has published a scientific paper within the two years preceding the patent's filing. However, the results lose significance when we define *active_inventor_author* based on any prior publication, regardless of timing. This supports the notion that corporate inventors must be actively engaged in scientific research at the time of patenting to benefit from network ties to academic scientists. As illustrated later, we also find consistent results when using the count measure of prior publications instead of a binary indicator.
[6] For ease of interpretation, our main specification uses a binary indicator. As illustrated later, we find consistent results when using the raw similarity measure between any of the corporate inventors' prior scientific papers and each of the twin papers, instead of a binary indicator.
[7] Alternatively, one might argue that the effect of social distance varies with the substantive relevance of the cited paper to the patent. To assess this, we computed embedding vectors for every patent and twin paper and calculated the cosine similarity between them. We then split the sample at the median similarity and re-estimated our main model separately for patents that are more versus less similar to the twin papers. Consistent with the idea that interpersonal ties matter more when the scientific prior art is more peripheral, the estimated effect of social distance is larger in the low-similarity subsample, although the difference is not statistically significant at conventional levels.

patent filing year), *patent stock* (number of unique patents by any of the paper's authors up to the patent filing), *institutional prestige* (number of unique PubMed papers linked to individuals with the same affiliations as the authors on the focal paper up to the year of patent filing), *U.S. based* (binary indicator equal to 1 if any affiliation of the authors on the focal paper is in the U.S.), *journal impact factor* (journal impact factor of the paper in year of publication), *journal commercial impact factor* (journal commercial impact factor of the paper in year of publication, available from Bikard and Marx (2020)), *paper citations* (number of times the focal paper is cited by other papers up to the year before patent filing), *patent citations* (number of times the focal paper is cited by patents up to the year before patent filing), and *published first* (binary indicator equal to 1 if the focal paper is published earlier or on the same day as the twin paper(s)). Prior work illustrates that scientific papers with more details and clarity are more likely cited in corporate patents (Bikard 2018). Therefore, we control for both *text length* (number of words in title and abstract) and *readability* (Gunning-Fog Index, calculated by considering the average number of words per sentence and the proportion of complex words with three or more syllables) (Hengel 2022). The control variable *patent citations* captures citations received by the focal paper from prior patents and therefore helps account for remaining unobserved paper-level heterogeneity that may affect the likelihood that the paper is cited by a patent.[8]

In addition to paper-level control variables, X_{ik} represents a vector of control variables at the paper-patent dyad level. We control for *time lag* as the number of days between the publication date of the paper and the filing date of patent, and for *min geographic distance* as the minimum distance in miles between any of the author(s) of the paper and any of the inventor(s) of the patent.

Finally, γ_{jk} accounts for fixed effects at the paper twin-patent level, effectively controlling for any unobserved heterogeneity present either at the level of the paper twin or the patent. This inclusion is crucial as it controls for several factors, including the nature and commercial potential of the scientific discovery, the specifics of the patented technology, as well as the characteristics of the company and the team of inventors behind the patent. Table 1 presents a detailed overview and definitions of all the variables.

'Insert Table 1'

To illustrate our empirical approach, consider the case of the CD40 academic twin discovery, which uncovered how CD40 signaling activates the immune system—a critical scientific insight that enabled the development of vaccines and immunotherapies targeting immune deficiencies. Building on this scientific discovery, Pfizer and Abgenix patented antibodies designed to activate CD40 to enhance immune responses (see US7288251, titled "Antibodies to CD40"). Nevertheless, the patent cites only

[8] Nevertheless, all findings remain consistent if we estimate the regressions without controlling for *paper citations* and *patent citations* (results available from the authors).

Bennett et al. (1998) and not Schoenberger et al. (1998), who made essentially the same scientific discovery at the same time.

'Insert Figure 1'

Figure 1 visualizes the social network connecting the corporate inventors named on the patent with the academic authors of both the cited and uncited twin papers. Each node represents an individual: red triangles indicate those who only publish, yellow dots those who only patent, and green squares those who do both. Corporate inventors on the citing patent are denoted by black light bulb icons; academic authors of the cited paper are marked with checked drawing board symbols, while academic authors of the uncited paper are marked with unchecked drawing board symbols. Edges represent co-authorship and/or co-inventorship ties up to 2002, the year the patent was filed. As shown in the figure, Ronald P. Gladue, a corporate inventor at Pfizer, had a prior co-authorship with Richard A. Flavell, an immunologist at Yale University and co-author of the cited Bennett et al. paper. This direct interpersonal tie (a shortest path of one) likely facilitated the transfer of knowledge from academia into corporate innovation. In contrast, the authors of the uncited twin paper by Schoenberger et al. (1998) were only indirectly connected to the corporate inventors, with a shortest path of two. Given that both papers reported essentially the same scientific discovery, the fact that only Bennett et al. (1998) was cited suggests that social proximity—rather than differences in content or commercial potential—may have influenced whether corporate inventors became aware of and built on a particular academic discovery.

4. Results

4.1 Summary statistics

Table 2 presents descriptive statistics comparing academic papers cited by company patents with their uncited twin counterparts. The analysis reveals no significant differences in team characteristics. The only notable exception is that cited papers tend to have authors with more prior patents, suggesting corporate inventors may pay more attention to research from academic scientists with stronger patenting track records. Similarly, cited and uncited papers do not differ significantly in paper characteristics.

'Insert Table 2'

Table 3 presents descriptive statistics at the paper-patent dyad level, comparing cited and uncited dyads. As expected, corporate inventors are socially closer to the academic authors of cited papers than to those of uncited ones: the average social distance is 2.75 for cited dyads versus 2.83 for uncited. Cited dyads are significantly more likely to have a shortest path of 0 or 1, indicating a direct or former direct collaboration tie. Notably, 2.8% of cited dyads exhibit a social distance of 0, in stark contrast to only 0.4% in uncited dyads. A zero-distance link reflects a direct connection where at least one corporate inventor is also an author of the cited academic paper. This overlap likely results from career mobility, as many PhD-

trained academic scientists in life sciences transition into industry (Sauermann and Roach 2012), especially given the average twelve-year gap between paper publication and patent filing. Alternatively, these individuals may retain academic positions while contributing to industrial R&D through consulting, contract research, or part-time roles. An additional 6.3% of cited dyads involve a prior co-authorship (distance = 1), versus 4.3% of uncited dyads—a pattern illustrated in Figure 1 by the connection between Ron Gladue (Pfizer) and Richard Flavell (Yale). In contrast, uncited dyads are more often connected at a distance of two or greater, though these differences are not always statistically significant. Overall, these findings suggest that closer interpersonal ties between corporate inventors and academic scientists increase the likelihood of knowledge transfer from academia to industry.

Prior research has primarily focused on direct collaborations between corporate inventors and academic scientists. In our cited paper–patent dyads, however, current or former direct ties account for only a small share of citations: approximately 3% occur at social distance 0 and 6% at social distance 1. Indirect ties are far more common: 26% of citations occur at social distance 2 and 53% at social distance 3. Although we cannot assess the total amount of knowledge transferred through each type of tie, these descriptive patterns suggest that a substantial share of observed academic knowledge flows to industry occurs through indirect interpersonal connections. Such indirect ties may expose corporate inventors to a broader and more diverse set of academic insights than the naturally restricted set of scientific discoveries produced by current or former collaborators. We confirm the external validity of this pattern using a random sample of corporate patent-to-academic paper citations to non-twin papers.[9]

'Insert Table 3'

4.2 Main effect of interpersonal connections

Table 4 presents our main results using binary indicators for social distance, with *social_dist_0* as the reference category. Model 4-1 includes controls only and Model 4-2 tests our main hypothesis. Consistent with *Hypothesis 1*, we find that corporate technology is significantly more likely to build on academic science when inventors are more closely connected to the academic authors through shorter interpersonal ties. The social distance indicators are jointly significant at $p<0.01$ (chi2(5)=36.10).

'Insert Table 4'

Because the coefficients in Table 4 are reported in log-odds units and are not directly interpretable as changes in citation probability, we translate them into predicted citation probabilities. Using the method of King et al. (2000), we estimate predicted probabilities and 95% confidence intervals at each level of

[9] We selected a random sample of 10,000 patent-to-paper citations from Marx and Fuegi (2020). As in our main analysis, we calculated *social_dist* as the shortest path between any inventor on the citing corporate patent and any academic author of the cited paper. The distribution of *social_dist* in this random sample is similar to the distribution observed for corporate patent citations to academic twin papers. Direct and former direct connections account for only a small fraction of citations: 3% occur at social distance 0 and 6% at social distance 1, closely matching the shares in our main sample. By contrast, indirect connections are far more common: 22% of citations occur at social distance 2 and 43% at social distance 3.

social distance. Figure 2 shows the results. When social distance is 0—that is, when an inventor is also a co-author of the focal academic paper—the predicted probability of citation is 68%. This probability falls to 52% at social distance 1, where an inventor and an academic author are prior direct collaborators, corresponding to a 16-percentage-point decrease, or a 24% relative decrease, compared with social distance 0. The predicted probability further declines to 43% at social distance 2, where an inventor and an academic author are connected through a common collaborator, corresponding to a 25-percentage-point decrease, or a 37% relative decrease, compared with social distance 0. At social distances of 3, 4, and 5 or more, the predicted probability of citation declines further to 38%, 38%, and 34%, respectively. These estimates reveal a clear monotonic decline in citation probability as social distance increases, consistent with the role of close interpersonal ties in facilitating knowledge transfer from academia to industry.

'Insert Figure 2'

Furthermore, the binary indicators for social distances of 0, 1, and 2 (*social_dist_0*, *social_dist_1*, and *social_dist_2*) are each statistically distinct from one another and from all higher-distance indicators (*social_dist_3*, *social_dist_4*, and *social_dist_5_or_more*). In contrast, the indicators for social distances of 3, 4, and 5 or more do not differ significantly from one another. These results point to a threshold in the effectiveness of interpersonal networks: while indirect ties up to two degrees of separation facilitate knowledge transfer, their influence becomes insignificant beyond that point.

'Insert Table 5'

Table 5 replicates Table 4 but replaces the binary indicators for social distance with a count measure, *social_dist*. This specification estimates the effect of moving one degree closer in the collaboration network between corporate inventors and academic authors. Across all models, *social_dist* is statistically significant at $p<0.01$. Because the coefficients are reported in log-odds units, we again translate them into changes in predicted citation probability. A one-unit decrease in social distance—representing one fewer intermediary between the corporate inventors and the academic authors—is associated with an 8- to 13-percentage-point increase in the predicted probability of citation (Models 5-2 to 5-8). The bottom row of Table 5 reports the observed citation probability in each column's estimation sample, calculated as the mean of the *cited* indicator. Because this probability is approximately 50% in these models, the marginal effects correspond to relative increases of approximately 16% to 26%.

Table A.3 in the Appendix presents results by firm type. Patents are categorized by assignee using data from Ewens and Marx (2024), including U.S. public firms, startups (3, 5, or 10 years old or younger at the time of patent filing), and firms with above- or below-median patent stocks (median = 560). Across all firm types, *social_dist* remains statistically significant at the 10% level or better, with no substantial

variation in effect size.[10] However, the effect appears somewhat stronger for startups than for older firms. Science-based startups often originate in academia, with many, especially in biotechnology and pharmaceuticals, founded by former university scientists or emerging directly from academic laboratories (Zucker et al. 1998). These startups typically rely heavily on external academic knowledge, making interpersonal networks particularly valuable for accessing and integrating frontier scientific insights.[11]

4.3 Endogenous formation of interpersonal connections

In line with prior work, we do not model the endogenous formation of collaborative teams, treating the resulting interpersonal network ties as exogenous (e.g., Singh 2005, Bikard et al. 2019, Marx and Hsu 2022).[12] [13] This assumption could lead to an overestimation of causal effects, if corporate inventors and academics formed interpersonal ties with the expectation of knowledge transfer. However, in our data, the average social distance in patent-to-paper citations is three degrees, meaning corporate inventors typically cite an academic collaborator of their collaborator's collaborator. If ties were endogenously formed to enable knowledge transfer, we would expect much shorter average social distances—that is, inventors would form direct or very close indirect ties specifically to gain access to knowledge. This interpretation is further supported by our finding that a social distance of three does not significantly influence knowledge transfer. In addition, the average patent in our sample cites 87 academic papers authored by 325 unique academic scientists and is co-invented by only four individuals, making it highly unlikely that all citations result from intentional tie formation with these academic scientists.

Nevertheless, to test whether our results are driven by ties that may have formed specifically to enable knowledge transfer, we exclude observations involving the closest forms of interpersonal proximity between academic scientists and corporate inventors. First, we omit observations where at least one corporate inventor is also an author of the focal academic paper (*social_dist* = 0).[14] Reassuringly, results

[10] In some smaller subsamples, the limited number of observations where certain binary indicators for social distance (specifically *social_dist_0*, *social_dist_1,* and *social_dist_5_or_more*) equal one reduces the precision of estimates. To address this, we use a count-based measure of social distance as the main specification.

[11] We also examined whether the effect of social proximity changed over time by interacting social distance with year dummies or five-year windows. Across all specifications, none of the interaction terms were statistically significant, suggesting no meaningful temporal change in the effect of social proximity over our observation window.

[12] Please note that other streams of research have modeled the formation of collaborative teams (Boudreau et al. 2017, Vakili et al. 2022).

[13] One possible alternative explanation is that inventors and academic authors share common training backgrounds, which could make the focal paper more likely to be cited independently of interpersonal ties. As a check against this explanation, we construct a proxy for shared PhD advisor or principal investigator. In the life sciences, the last author on early-career papers typically represents the PI or lab head. We identify early-career mentors as the last authors on papers published during the first five years of each inventor's or author's publishing career and define *shared PI* as equal to one if any inventor on the patent and any author of the focal twin paper share such a mentor. Adding this control does not materially change any of the findings. Results are available from the authors.

[14] Collecting data on patent–paper pairs from Marx (2026), we find that fewer than 1% of all paper–patent dyads in our sample represent such pairs. All findings remain consistent after excluding these observations from the analysis (results available upon request from the authors).

remain consistent (Table 4, Model 4-3; Table 5, Model 5-3). The findings also hold when excluding both same-individual overlap and prior direct collaboration ties (*social_dist* = 0 or 1; Table 4, Model 4-4; Table 5, Model 5-4). As expected, the effect size of *social_dist* decreases when observations with the shortest social distances are excluded but remains large and statistically significant ($p<0.01$). A one-unit decrease in social distance increases the predicted probability of citation by approximately 13 percentage points in the full model (Model 5-2), 12 percentage points when excluding observations with social distance 0 (Model 5-3), and 8 percentage points when excluding observations with social distances 0 and 1 (Model 5-4). Given that the observed citation probability is approximately 50% in these estimation samples, these marginal effects correspond to relative increases of approximately 26%, 24%, and 16%, respectively.

As an additional check, we recalculate social distance using only co-authorship ties formed before the publication of the twin papers, excluding all ties created during the average 12-year lag between paper publication and patent filing. This approach, by design, eliminates interpersonal ties formed after the academic twin discovery—ties that might have been intended to facilitate knowledge transfer and support the commercial exploitation of the scientific discovery. The effect remains strong, statistically significant at $p<0.01$, and similar in magnitude (Table 5, Model 5-6). A one-unit decrease in *social_dist* increases the predicted probability of citation by 11 percentage points when using only pre-publication ties, compared with 13 percentage points in the standard specification, which incorporates interpersonal ties up to the patent filing year. Given that the observed citation probability is approximately 50% in these estimation samples, these marginal effects correspond to relative increases of approximately 22% and 26%, respectively. Due to the 12-year average lag between paper publication and patent filing, 30% of paper–patent dyads involve inventors or authors with no publications or patents prior to the twin papers' publication, and hence result in missing values for *social_dist* in this subsample.[15] The limited number of non-zero observations for certain binary indicators of social distance makes precise estimation difficult. To address this, we use the count-based measure of *social_dist* as the primary specification (Table 5, Model 5-6). As shown in Table 4, Model 4-6, our results remain robust when using binary indicators for social distance, with the exception that *social_dist_1* is no longer statistically distinguishable from *social_dist_2* and *social_dist_3* at conventional levels—unsurprising given the small number of dyads at distance 1 when restricting to co-authorship links formed before the twin papers' publication.

As a final check for endogenous tie formation, we manually identify and exclude twin papers that acknowledge any form of corporate support or funding, including direct financial support, provision of research materials or tools, or acknowledged support from individuals with a corporate affiliation. The

[15] Consistent with our approach to handling missing values, we impute these as the maximum observed social distance plus one (i.e., 9). Reassuringly, when excluding these observations entirely from the analysis, the findings are nearly identical: a one-unit decrease in *social_dist* is associated with a 12.72 percentage point increase in the probability of citation when imputing, compared to 12.63 percentage points when dropping the observations from the analysis.

results remain unchanged (Table 4, Model 4-7; Table 5, Model 5-7). Together, these checks suggest that our findings are unlikely to be driven by the intentional formation of interpersonal ties to enable knowledge transfer from academia to industry.

4.4 Corporate inventors' scientific expertise and active engagement in scientific research

Having established the role of social proximity in enabling corporate inventors to build on academic science, we next examine whether inventors' absorptive capacity conditions this effect. Specifically, we test whether active engagement in scientific research and alignment of scientific expertise strengthen the impact of social proximity on the use of academic science.

To test *Hypothesis 2*, we extend Model 5-2 by including an interaction term (*social_dist* × *active_inventor_author*). Note that the main effect of *active_inventor_author* is absorbed by the patent-level fixed effects and cannot be separately estimated. As shown in Table 6, Model 6-2, the interaction term is highly significant ($p<0.01$), while the main effect of *social_dist* becomes insignificant and drops sharply (from –0.957 to –0.068). For corporate teams without inventors active in scientific research, a one-unit decrease in *social_dist* is associated with a statistically insignificant 1-percentage-point increase in the predicted probability of citation. In contrast, for teams with active inventor-authors, the same one-unit decrease in *social_dist* increases the predicted probability of citation by 15 percentage points, an effect that is statistically significant at $p<0.01$. Given that the observed citation probability is approximately 50% in this estimation sample, this corresponds to a relative increase of approximately 30%. Table A.4 in the appendix replicates all robustness checks from Table 5, confirming the stability of the interaction effect between *social_dist* and *active_inventor_author*.[16] As an additional robustness check, column 11 in Table A.4 uses a more granular count measure rather than a binary indicator, calculating *active_inventor_author* as the number of unique scientific papers published by any co-inventor of a patent in the year preceding the patent's filing. Together, these findings indicate that interpersonal ties to academic scientists are valuable primarily when corporate inventors are actively engaged in scientific research. Absent such active engagement, social proximity by itself does not appear sufficient to translate academic insights into corporate technology.

'Insert Table 6'

To test whether closely related scientific expertise enables corporate inventors to more effectively leverage interpersonal connections with academic scientists (*Hypothesis 3*), we introduce an interaction term (*social_dist* × *close_paper*) in Model 6-3 of Table 6. While the main effect of *close_paper* is absorbed by the patent-level fixed effects, its interaction with *social_dist* is highly significant ($p<0.01$).

[16] The only exception is that the interaction effect *social_dist_time* × *active_inventor_author* is observed at p = 0.104 in Column 10, falling slightly outside conventional significance thresholds.

Including this interaction substantially reduces the magnitude of the *social_dist* coefficient (from -0.957 to -0.376). Among inventor teams lacking closely related scientific expertise (*close_paper* = 0), a one-unit decrease in social distance increases the predicted probability of citation by 5 percentage points. For inventor teams with closely related scientific expertise (*close_paper* = 1), the same one-unit decrease increases the predicted probability of citation by 17 percentage points. Given that the observed citation probability is approximately 50% in this estimation sample, these marginal effects correspond to relative increases of approximately 10% and 35%, respectively. Table A.5 replicates all robustness checks from Table 5, confirming the stability of this interaction effect. As an additional check, column 11 in Table A.5 uses a more granular continuous measure rather than a binary indicator, calculating *close_paper* as the maximum cosine similarity between any of inventors' papers published before the patent filing year and any of the twin papers. These findings underscore that the ability to absorb external academic knowledge through interpersonal connections is strongly shaped by the inventors' own scientific expertise and background knowledge.

In Model 6-4, which includes both interaction terms, both remain statistically significant ($p<0.01$ and $p<0.05$), while the main effect of *social_dist* becomes insignificant. The three terms are jointly significant at $p<0.01$. Taken together, these results indicate that corporate inventors benefit most from interpersonal ties to academia when they are actively engaged in scientific research and possess closely related scientific expertise. These findings suggest that social proximity and inventors' absorptive capacity operate jointly, enabling inventors to access, interpret, and integrate external scientific knowledge into technology development.

4.5 Comparing the effect of interpersonal connections on science-to-science versus science-to-technology knowledge transfer

As theorized and empirically demonstrated in this paper, accessing and integrating scientific findings into corporate technology depends on close interpersonal connections between corporate inventors and the academic scientists who produced those scientific discoveries. A remaining question, however, is whether interpersonal proximity operates differently for science-to-technology knowledge transfer (our focal context) than for science-to-science knowledge transfer (Balsmeier et al. 2025). To examine this, we replicate our empirical strategy using paper-to-paper rather than patent-to-paper citations to the same set of twin papers. We construct a dataset of PubMed articles published between 1975 and 2009 that cite exactly one (but not all) of the papers in a twin. Following the same identification approach as in the patent-to-paper analysis, each citing paper is paired with both the cited and the uncited twin paper, yielding a matched set of cited and uncited paper–paper dyads. For each dyad, we compute the shortest path between the authors of the citing paper and the authors of respectively the cited and uncited twin

paper using the same integrated author–inventor collaboration network as before. The resulting dataset contains 221 paper twins (448 twin papers), 21,571 citing papers, and 44,220 paper–paper dyads, including 22,069 cited dyads and 22,151 uncited dyads.

Table A.6 presents descriptive statistics at the paper–paper dyad level, comparing cited and uncited dyads, parallel to the descriptive statistics for patent-paper dyads in Table 3. Consistent with the patent-to-paper patterns, authors of citing papers are socially closer to the authors of the cited twin paper than to the authors of the uncited one, indicating that interpersonal proximity also shapes science-to-science knowledge transfer. As expected, authors of citing papers are typically more closely connected to the authors of the cited paper than citing corporate inventors are to the cited authors. In the paper-to-paper data, 59.2% of dyads have a social distance of two or less, whereas in the patent-to-paper setting only 34.8% of dyads fall within the same range. This difference underscores that interpersonal proximity is naturally stronger within the academic community than across the boundary between academia and industry.

Table A.7 replicates the conditional logit analysis for paper-to-paper citations, using the same count measure of social distance as in the patent-to-paper analysis reported in Table 5. Across all models, *social_dist* is statistically significant at $p<0.01$. To compare effect sizes, we again translate the log-odds coefficients into changes in predicted citation probability. In the paper-to-paper analysis, a one-unit decrease in social distance—representing one fewer intermediary in the interpersonal network—is associated with a 9-percentage-point increase in the predicted probability of citation (Table A.7, Column 2). Given that the observed citation probability in this estimation sample is approximately 50%, this corresponds to a relative increase of approximately 17%. The corresponding effect in the patent-to-paper analysis is larger. A one-unit decrease in social distance is associated with a 13-percentage-point increase in the predicted probability of a patent-to-paper citation (Table 5, Column 2), corresponding to a relative increase of approximately 26% given the similar observed citation probability in that estimation sample. Because the mean observed citation probabilities are nearly identical across the two estimation samples, the percentage-point and relative effects can be compared on a common probability scale. A formal test of coefficient differences confirms that the effect of *social_dist* is statistically stronger for patent-to-paper than for paper-to-paper citations ($z = -3.08$, $p = 0.002$). The pseudo-R^2 is also lower for the paper-to-paper model than for the patent-to-paper model (0.112 versus 0.396), consistent with interpersonal proximity playing a smaller role in explaining science-to-science citations than science-to-technology citations. Figure A.2 provides the direct analogue to Figure 2 for the paper-to-paper citation analysis, replacing patent-to-paper citations with citations from later scientific papers to the focal twin papers. As in Figure 2, the estimates are based on binary indicators for social distance and are reported as predicted citation probabilities. For paper-to-paper citations, the predicted probability of citation is 56% at social distance 0,

41% at social distance 1, 34% at social distance 2, 32% at social distance 3, and 30% at social distances 4 and 5 or more. The corresponding predicted probabilities for patent-to-paper citations in Figure 2 are 68%, 52%, 43%, 38%, 38%, and 34%.

Taken together, these results show that interpersonal proximity matters in both science-to-science and science-to-technology knowledge flows, but that its role is more pronounced in the latter. One interpretation is that translating academic science into corporate technology requires crossing greater cognitive and institutional boundaries, making close interpersonal ties especially valuable for recognizing, interpreting, and applying academic insights. A complementary interpretation is that academic scientists have more alternative channels for learning about relevant discoveries, such as conferences, seminars, editorial roles, and overlapping academic communities, which may partly substitute for social proximity. Our evidence cannot distinguish between these mechanisms, but it does show that interpersonal proximity plays a larger role when academic science is translated into corporate technology than when scientific knowledge circulates within academia.

4.6 Robustness checks

Our findings are robust for the subsample of patent to paper citations that appear in the body text of the patent, excluding those citations only listed on the front page (Table 4, Model 4-8; Table 5, Model 5-8). In-text citations are more likely included by the inventors themselves, perhaps providing a more accurate insight into academic knowledge transfers to corporate inventors (Bryan et al. 2020). Our findings also remain robust when we expand the sample to include all instances where a corporate patent cites every paper from the same twin discovery, rather than only a subset as in our previous analysis (Table 4, Model 4-9; Table 5, Model 5-9).

While our measure of social distance does not account for tie decay, one could argue that more recent interpersonal ties facilitate knowledge transfer more effectively than older ones (Hansen 1999). To address this, we introduce a time-weighted measure of social distance, *soc_distance_time*, which prioritizes the shortest paths between academic authors and company inventors while incorporating the recency of their collaborations (Wasserman and Faust 1994). This approach ensures that more recent collaborations contribute less to the overall path length compared to older ties, thereby emphasizing the temporal relevance of connections. The exact calculation is detailed in Table 1. As shown in model 5-10 of Table 5, our findings remain robust. Figure A.3 displays the predicted probability of citations for different values of *soc_distance_time.*

‘Insert Table 7’

Table 7 presents additional robustness checks related to the criteria used to identify paper twins and the sample of corporate patents included in the analysis. In certain smaller subsamples, the small

number of observations where some binary indicators for social distance equal one makes obtaining accurate estimates challenging. Consequently, we utilize the count measure for social distance as our main specification. Nevertheless, we find qualitatively similar results using binary indicators for social distance as shown in Table A.8.[17] Model 7-1 excludes observations linked to higher order twins, i.e., triplets. Model 7-2 is confined to twin papers published within a three-month instead of a one-year window. Model 7-3 includes only twin papers with a very high degree of content similarity, defined as having a cosine similarity score above 0.9. In Model 7-4, we focus on twin papers that each rank first on PubMed's list of most similar papers to one another. Model 7-5 runs the analysis based on the subset of twin papers that are published back-to-back in the same journal, volume and issue. Model 7-6 selects twin papers cited adjacently 100% of the time, indicating they are always mentioned together in the references of other papers. Model 7-7 excludes papers authored by large teams, specifically those with more than five authors. Model 7-8 removes papers that were highly cited before the patent filing, specifically those with over 1,000 citations. Model 7-9 disregards observations linked to the two companies most frequently listed as patent assignees in our dataset, Genentech and Millennium Pharmaceuticals. Despite substantial differences in the composition of the subsamples, the results remain consistent: *social_dist* is statistically significant at $p<0.01$ in all models, and a one-unit decrease in social distance is associated with a 12- to 17-percentage-point increase in the predicted probability of citation. Given that the observed citation probability is approximately 50% across these estimation samples, these marginal effects correspond to relative increases of approximately 24% to 34%.

Because interpersonal and geographic proximity are correlated, we verify that our findings are not driven by the inclusion of the geographic distance control. Tables A.9 and A.10 report the main models with and without controlling for *min geographic distance*. The results remain robust when *min geographic distance* is excluded. Thus, geographic proximity does not account for the relationship between interpersonal proximity and citation likelihood. If anything, controlling for geographic distance makes the estimated relationship between social distance and probability of citation stronger in magnitude.

4.7 Characterizing corporate inventors with ties to academia

Our empirical identification strategy estimates the effect of social proximity while controlling for unobserved heterogeneity at the patent level. The inclusion of patent fixed effects absorbs any time-invariant characteristics of the patented technology, the firm, and the team of corporate inventors. Even so, better understanding who the corporate inventors are that most directly bridge academic science and corporate technology remains important. Building on the above results, we therefore examine the

[17] Please notice that estimations for the subsamples of twins with PubMed rank order=1 or published back-to-back did not converge because of insufficient observations for some of the different binary categories of social distance.

characteristics and career trajectories of two groups of inventors: (i) those on corporate patents who hold the closest interpersonal ties to the academic authors of the cited paper, and (ii) those who actively engaged in scientific research (active inventor–authors). Using the full publication and patent histories of all corporate inventors,[18] we reconstruct prior affiliations, scientific and inventive output, and broader career trajectories for each inventor.

Table A.11 in the Appendix reveals clear differences between inventors with the shortest interpersonal ties to cited academic authors and other inventors on the same corporate patents without shortest ties. Socially proximate inventors are substantially more likely to have past and recent affiliations with universities or public research institutes, including shared affiliations with the authors of the cited papers. They also exhibit stronger scientific engagement: they began publishing earlier, have larger cumulative publication records, and remain more active in recent scientific publishing, and show higher recent patenting activity. Finally, they have a more extensive collaboration network, with a substantially larger number of inventors, authors, and inventor–authors within three degrees of separation. Taken together, these patterns indicate that socially proximate inventors are those with more recent and overlapping academic affiliations, longer scientific publishing experience, greater recent scientific and inventive activity, and markedly larger collaboration networks—features that position them especially well to bridge academic science and corporate technology.

In parallel, Table A.12 in the Appendix compares corporate inventors who were actively publishing in the year before patent filing with those who were not. Actively publishing inventors display a distinct profile: they are substantially more likely to have recent and past affiliations with universities or public research institutes—including shared affiliations with the cited academic authors—and they exhibit markedly higher scientific and inventive output. They also tend to have significantly larger collaboration networks. Taken together, these patterns indicate that inventors who continue to publish while working in industry remain more academically embedded and have broader collaborative reach.

5. Discussion and Conclusion

Scientific research fuels technological progress in industry, yet essential scientific knowledge increasingly originates outside the firm, in universities and public research labs. This shift raises a critical question: how can firms identify, interpret, and apply external academic knowledge? Our study highlights how

[18] Affiliation data from publications and patents are imperfect and may contain noise or omissions, especially for older records. In scientific publications, such errors are more common because papers often involve many co-authors with multiple affiliations, and automated parsing can incorrectly assign institutional affiliations to specific authors. One alternative would be to use LinkedIn to trace inventors' career histories, but this source is incomplete for our study period. In a manual check of 20 randomly selected inventors, 30% had no LinkedIn profile, and 35% of those with a profile omitted key prior affiliations or positions that we could verify through other sources. As a result, LinkedIn does not provide sufficiently comprehensive or reliable coverage for our purposes. Our findings should therefore be interpreted with appropriate caution regarding likely measurement noise in affiliation histories.

interpersonal networks between corporate inventors and academic scientists combine with inventors' absorptive capacity to facilitate this process. By leveraging nearly identical scientific discoveries published by separate academic teams and examining how they are incorporated into corporate technology, we isolate the influence of interpersonal connections from the content and commercial potential of the scientific discoveries themselves.

Our findings indicate that firms draw on both direct and indirect connections to access scientific insights from academics. Direct ties are more effective at facilitating knowledge transfer, but a larger share of observed knowledge transfers occurs through the greater number of indirect connections. Consistent with network theory, we also observe limits to the reach of these networks. Beyond a social distance of two degrees—that is, beyond cases in which a corporate inventor and an academic scientist are connected through a mutual collaborator—the network effect dissipates. Moreover, consistent with individual-level absorptive capacity theory, inventors' ability to leverage social proximity to academics depends on their active engagement in scientific research and on the alignment between their scientific expertise and the academic knowledge they seek to integrate. Finally, replicating our design using paper-to-paper rather than patent-to-paper citations to the same twin papers, we find that interpersonal proximity also shapes science-to-science knowledge transfer, but less strongly. This comparison suggests that interpersonal proximity plays a larger role when scientific knowledge is translated into corporate technology than when it circulates within academia.

Our study faces several limitations. First, relying on patent citations to academic papers captures only a subset of the scientific knowledge relevant for technological development. Yet, surveys suggest such citations correlate with managers' reports of leveraging academic research (Roach and Cohen 2013) and reflect inventors' awareness of the papers (Bikard and Marx 2020). Unlike patent references, most citations to scientific papers are added by inventors rather than examiners, making them more reliable indicators of knowledge transfer (Marx and Fuegi 2020). Moreover, paper twins represent a non-random subset of academic work, likely capturing frontier discoveries with substantial scientific and commercial potential, often produced by prominent scientists and institutions. Also, we construct interpersonal networks using co-authorship links and advance this line of research by leveraging a new dataset that captures all co-authorships on both patents and papers from PubMed and U.S. patents, while also providing a structured crosswalk between the two. However, our approach potentially does not capture interpersonal ties formed through other channels, such as conferences, workshops, or industry meetings (Baruffaldi and Poege 2025).

Relatedly, while our analysis documents a strong association between interpersonal proximity and the use of academic science in corporate technology, our large-scale data cannot identify the precise micro-level processes through which such proximity operates. In the theory section, we argued that

interpersonal ties may facilitate science-to-technology transfer by raising awareness of relevant scientific advances, reducing search costs, providing access to contextual and partly tacit knowledge, building credibility and trust, enabling referrals, and offering interpretive support. Our empirical evidence is consistent with this broader set of mechanisms, but it does not distinguish among them directly. The findings should therefore be read as showing that interpersonal proximity facilitates the use of academic science in corporate technology, rather than as direct evidence of any single underlying mechanism. Future research could unpack these micro-level processes in greater depth.

Furthermore, our study focuses on the life sciences, where science–industry linkages are particularly strong. The extent to which our findings generalize to other fields therefore remains an open question. We expect interpersonal proximity to matter most when scientific knowledge is more tacit, complex, or rapidly evolving, when academic research is a key input to technological development, and when cross-sector mobility and collaboration are common. These conditions characterize much of the life sciences but are less typical in domains such as mechanical engineering, where R&D is more internally oriented and science–technology linkages are weaker (Cohen et al. 2002). The role of interpersonal networks should be more limited in settings where scientific knowledge is highly codified, standardized, or embedded in mature technologies. At the same time, the life sciences account for roughly one-third of global scientific output (National Science Board 2023) and a disproportionately large share of science-to-technology flows, underscoring the relevance of our empirical setting. Future comparisons across fields could further refine these boundary conditions.

Also, corporate inventors who possess greater skills or better recognize the potential benefits are perhaps more likely to form interpersonal ties that allow them to occupy valuable network positions and act as cross-institutional boundary-spanners within the social network (Lee 2010). Yet, our analysis controls for any unobserved heterogeneity at the patent level and thus for the corporate inventors on the patent. For academic scientists, our estimations control for various characteristics that might confound the influence of interpersonal connections on knowledge transfer, such as their stock of prior patents and papers, the size of their social network, and the prestige of their academic institution.

Finally, we lack exogenous variation in social proximity, corporate inventors' engagement in scientific research, and their scientific expertise. Several robustness checks—such as excluding direct or prior collaborations and measuring social proximity before the scientific discovery—suggest that endogenous tie formation is unlikely to drive our results. Nonetheless, as with any observational design, we cannot fully rule out all alternative explanations, and our findings should therefore be interpreted with this uncertainty in mind.

Within these limitations, our study advances theory and offers implications for practice. Adopting a micro-foundations perspective, we conceptualize firms' ability to benefit from academic science as an

emergent outcome of individual inventors' social connections, active engagement in scientific research, and closely related scientific expertise. Building on this view, we contribute to research on how firms benefit from academic science by illuminating the social dimension of absorptive capacity. Interpersonal ties to academics help firms leverage external scientific knowledge when corporate inventors have the scientific engagement and closely related expertise needed to recognize, interpret, and apply that knowledge in corporate innovation. In doing so, we bridge network theory, which highlights the role of interpersonal connections—especially indirect ties—in accessing novel and distributed knowledge, with absorptive capacity theory, which emphasizes how active engagement in scientific research and closely related scientific expertise enable individuals to recognize, assimilate, and apply external academic insights in corporate innovation.

While prior research has largely focused on direct interpersonal collaborations between firms and academia (Liebeskind et al. 1996, Cockburn and Henderson 1998, Zucker et al. 2002, Agrawal 2006), our study adopts a broader network perspective by analyzing the full interpersonal network structure that links corporate R&D staff to the academic community. Our findings highlight social proximity—up to two degrees of separation—as a key mechanism enabling the transfer of academic knowledge to industry. Notably, direct collaboration ties to academic scientists represent the most effective conduit for knowledge transfer, yet account for only a small fraction—approximately 9%—of all observed transfers. In contrast, indirect ties at a social distance of two, where corporate inventors are connected to the academic scientist who made the scientific discovery only via a common collaborator, still significantly support knowledge transfer and account for nearly 26% of all observed transfers. These results resonate with network theory on the value of indirect or weak ties in accessing non-redundant, novel, and distributed knowledge not available through one's immediate circle (Granovetter 1973, Cowan et al. 2007). At the same time, our finding that indirect ties beyond two degrees of separation do not facilitate knowledge transfer aligns with Hansen's (1999) insight that weak ties may lack the relational strength or informational bandwidth required to transmit complex or partly tacit scientific knowledge. Future work should go beyond citations as a measure of knowledge transfer, for example by using text-based approaches to estimate finer-grained measures of knowledge flows (Arts et al. 2025), thereby enabling more precise comparisons of the volume of knowledge transferred through direct versus indirect ties.

Our work also extends prior research showing that interpersonal networks among academic scientists shape the transfer of scientific knowledge within academia (Mohnen 2022), as well as studies demonstrating how networks of corporate inventors facilitate the transfer of technical knowledge within and across firms (Singh 2005, Paruchuri and Awate 2017, Moreira et al. 2018). Singh (2005), in particular, provides important evidence that patent-to-patent knowledge flows among inventors are more likely when inventors are connected through prior collaborative ties, and that this relationship weakens as

social distance increases. We examine a different knowledge-transfer setting: the movement of scientific knowledge across the boundary between academia and industry. Specifically, we focus on patent-to-paper citations and use an integrated collaboration network that spans academic scientists and corporate inventors. Our research also extends prior work that mapped collaboration networks linking academic scientists and corporate inventors (Balconi et al. 2004, Breschi and Catalini 2010, Lissoni 2010) by providing empirical evidence that these network configurations help explain the transfer of scientific knowledge from academia to industry.

Our findings also contribute to classic work on boundary-spanning individuals—those who serve as organizational gatekeepers by acquiring, translating, and integrating external knowledge (Allen 1977, Tushman 1977, Tushman and Scanlan 1981). In our context, corporate inventors who are simultaneously engaged in scientific research and commercial technology development fulfill this role. Prior research has shown that such dual-engaged Pasteur's Quadrant Researchers (PQRs; Scharfmann et al. 2025) enhance firm innovation outcomes (Cockburn and Henderson 1998, Zucker et al. 2002, Gittelman and Kogut 2003), yet the mechanisms behind this effect remain underexplored. Our study provides micro-level evidence that these individuals leverage their scientific expertise and academic ties to bridge the gap between scientific discovery and technological application. In line with theories of individual absorptive capacity (Ellis 1965, Cohen and Levinthal 1990) and boundary-spanning (Tushman and Scanlan 1981), we find that only those inventors with active engagement in scientific research and closely aligned scientific expertise are able to effectively draw on interpersonal connections to absorb and apply academic knowledge. This dual requirement—scientific proximity and social proximity—appears essential to translate external academic discoveries into corporate innovation.

For managers and policymakers, our findings carry several implications. One is the importance of cultivating human bridges to academia, not just through direct collaborations but also indirect ties. Firms should recruit and develop R&D personnel who are fluent in the language of science—for example, PhD-trained researchers who continue to publish or collaborate with universities. These individuals not only bring their own scientific expertise, but also have professional networks that can alert the firm to important scientific discoveries and help interpret their relevance for corporate technology. The firm's culture should support their continued engagement with the scientific community (e.g., attending conferences, publishing) that builds connections and absorptive capacity. Given the documented decline in firms' internal investments in scientific research and their reduced publication output (Arora et al. 2018), our findings underscore the risks of this trend. While reduced investment in scientific research may limit knowledge spillovers to competitors, it may inadvertently erode absorptive capacity, weaken interpersonal networks, and isolate the firm from the upstream scientific knowledge generation (Cockburn and Henderson 1998, Arora et al. 2021).

Second, our research challenges the notion of academic science as a purely public good that any firm can exploit equally without any restrictions or competition (Nelson 1959). In theory, once academic research is published, it's available. In practice, we find that firms with close social ties to the academics who made that scientific discovery gain a disproportionate advantage in using that research. This implies a form of natural exclusion—firms lacking those social connections may be left behind in leveraging science, which can lead to underutilization of academic science in industrial R&D. For managers, this underscores the value of investing in relationship-building with universities and participating in knowledge networks (e.g., industry consortia, research alliances, sponsoring conferences). For policymakers and university administrators, it suggests that simply publishing research is not enough to ensure broad dissemination (Mumtaz 2025). Facilitating industry–academia interactions (through joint workshops, researcher exchanges, or industry liaison programs) is vital so that valuable academic knowledge does not languish unused. Encouraging informal social networks, such as alumni networks and conference communities, can pay dividends by spreading know-how to where it can be applied. Our study focused on co-authorship ties as a measurable subset of relationships, but the benefits we observe likely extend to other forms of professional connections and social capital (Baruffaldi and Poege 2025). Policies that encourage scientists to engage with industry (for instance, by valuing applied collaborations in academic promotion, or providing grants that require industry partners) could enhance these knowledge flows.

Lastly, our findings re-emphasize the importance of geographical proximity and cluster-based innovation policy in a new light. It has long been argued that colocation of firms near universities boosts knowledge spillovers (Zucker et al. 1998, Alcácer and Chung 2007). Our results provide a micro-mechanism for this: physical proximity fosters interpersonal ties between academics and corporate inventors—through chance meetings, local networks, repeated interactions—that in turn facilitate the transfer of scientific knowledge to industry. Thus, geographic clusters create fertile ground for the very social networks between academia and industry we find so critical. At the same time, social connections themselves may also shape where individuals and organizations choose to locate, suggesting a mutually reinforcing dynamic between geographic and social proximity. For both firms and policymakers, this means that investments in regional innovation hubs (incubators, joint research centers) are not just about infrastructure or formal partnerships, but about enabling the dense web of social connections through which knowledge truly flows.

References

Agrawal A (2006) Engaging the inventor: Exploring licensing strategies for university inventions and the role of latent knowledge. *Strategic Management J.* 27(1):63–79.

Alcácer J, Chung W (2007) Location strategies and knowledge spillovers. *Management Sci.* 53(5):760-776.

Allen T (1977) *Managing the flow of technology: the transfer and the dissemination of technological information within the R&D organization* (MIT Press, Cambridge, MA).

Anckaert PE, Cassiman B (2026) Fostering creativity through the exploitation of scientific and technological knowledge: An in-depth study of technology development in the lithium-ion battery field. *Ind. Corporate Change* 35(1):1–33.

Arora A, Belenzon S, Patacconi A (2018) The decline of science in corporate R&D. *Strategic Management J.* 39(1):3–32.

Arora A, Belenzon S, Sheer L (2021) Knowledge spillovers and corporate investment in scientific research. *Amer. Econom. Rev.* 111(3):871-898.

Arts S, Veugelers R (2020) Taste for science, academic boundary spanning, and inventive performance of scientists and engineers in industry. *Indust. Corporate Change* 29(4):917-933.

Arts S, Melluso N, Veugelers R (2025) Beyond citations: Measuring novel scientific ideas and their impact in publication text. *Rev. Econom. Statist.*, forthcoming. https://doi.org/10.1162/rest_a_01561

Azoulay P, Ganguli I, Graff Zivin J (2017) The mobility of elite life scientists: Professional and personal determinants. *Res. Policy* 46(3):573–590.

Balconi M, Breschi S, Lissoni F (2004) Networks of inventors and the role of academia: an exploration of Italian patent data. *Res. Policy* 33(1):127-145.

Balsmeier B, Lück S, Fleming L (2025) Science knowledge localizes. *Res. Policy* 54(10):105333.

Baruffaldi SH, Poege F (2025) Like Stars: How Firms Learn at Scientific Conferences. *Management Sci.* 71(3):1865-1888.

Bennett S, Carbone F, Karamalis F, Flavell R, Miller J, Heath W (1998) Help for cytotoxic-T-cell responses is mediated by CD40 signalling. *Nature* 393(6684):478–480.

Bikard M (2018) Made in academia: The effect of institutional origin on inventors' attention to science. *Organ. Sci.* 29(5):818–836.

Bikard M (2020) Idea twins: Simultaneous discoveries as a research tool. *Strategic Management J.* 41(8):1528–1543.

Bikard M, Vakili K, Teodoridis F (2019) When collaboration bridges institutions: The impact of university–industry collaboration on academic productivity. *Organ. Sci.* 30(2): 426-445.

Bikard M, Marx M (2020) Bridging academia and industry: How geographic hubs connect university science and corporate technology. *Management Sci.* 66(8):3425–3443.

Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg A, Lakhani KR (2017) A field experiment on search costs and the formation of scientific collaborations. *Rev. Econom. Statist.* 99(4):565-576.

Breschi S, Catalini C (2010) Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. *Res. Policy* 39(1):14–26.

Bryan KA, Ozcan Y, Sampat B (2020) In-text patent citations: A user's guide. *Res. Policy* 49(4):103946.

Bush V (1945) *Science - The Endless Frontier*. A Report to the President on a Program for Postwar Scientific Research. National Science Foundation, Washington, DC.

Cavallo A, Colombo MG, Foss NJ (2025) Internal organization and absorptive capacity in open innovation contexts: The roles of gatekeeping, task specialization, and delegation. *Indust. Corporate Change* 34(5):1006-1028.

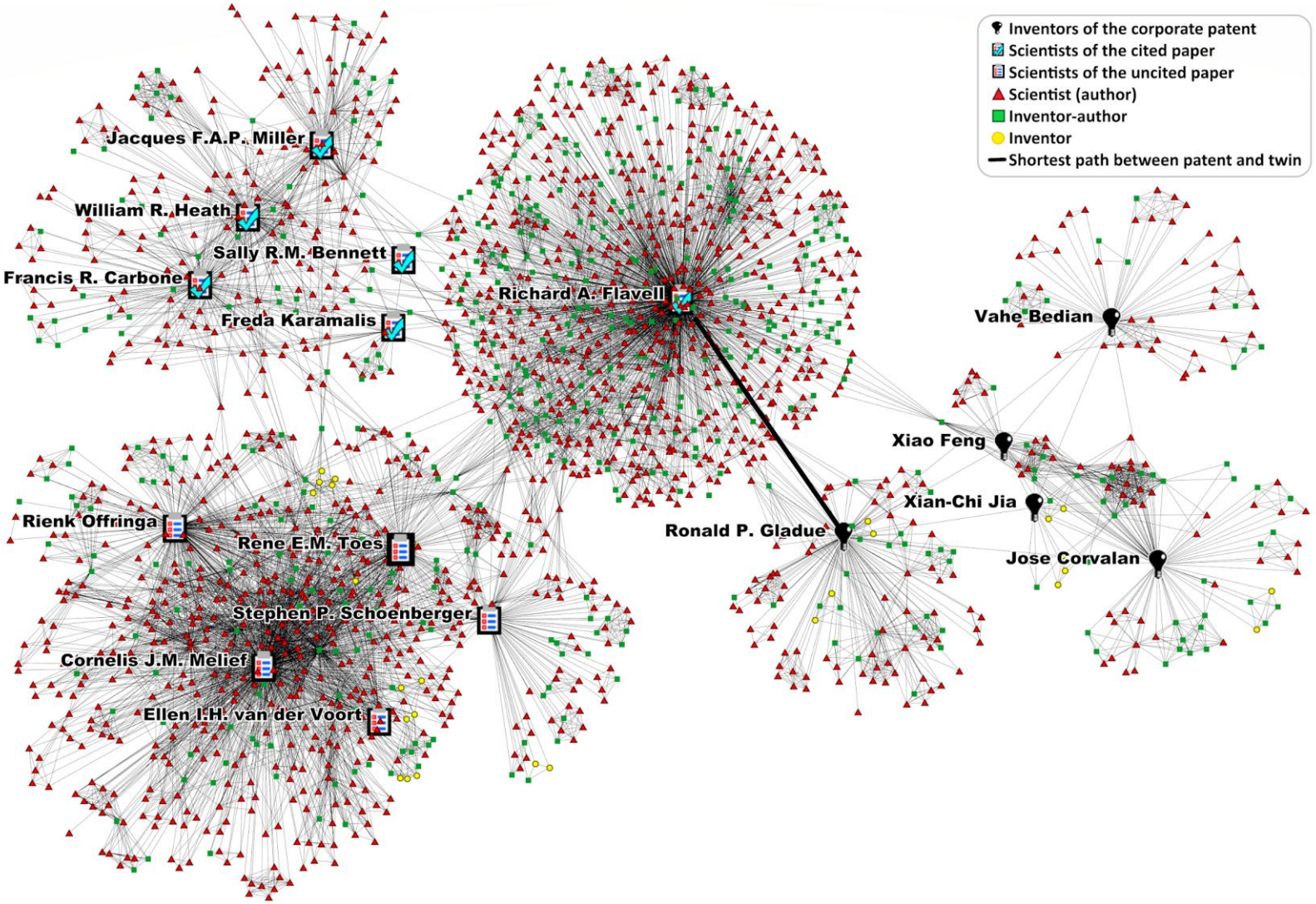
Cockburn IM, Henderson RM (1998) Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *J. Indust. Econom.* 46(2):157–182.

Cohan A, Feldman S, Beltagy I, Downey D, Weld DS (2020) SPECTER: Document-level representation learning using citation-informed transformers. *Proc. 58th Annu. Meeting Assoc. Comput. Linguist. (ACL 2020)* Online*:*2270–2282.
Cohen WM, Levinthal DA (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Admin. Sci. Quart.* 35(1):128–152.
Cohen WM, Nelson RR, Walsh JP (2002) Links and impacts: the influence of public research on industrial R&D. *Management Sci.* 48(1):1-23.
Collins HM (1974) The TEA Set: Tacit Knowledge and Scientific Networks. *Sci. Stud.* 4(2):165–185.
Cowan R, Jonard N, Zimmermann JB (2007) Bilateral collaboration and the emergence of innovation networks. *Management Sci. 53*(7):1051-1067.
Ellis HC (1965) *The Transfer of Learning* (Macmillan, New York).
Ewens M, Marx M (2024). Firm Age and Invention: An Open-Access Dataset. Working Paper, Columbia Business School, New York, NY and Cornell University, Ithaca, NY.
Felin T, Foss NJ, Heimeriks KH, Madsen TL (2012) Microfoundations of routines and capabilities: Individuals, processes, and structure. *J. Management Stud.* 49(8):1351-1374.
Fleming L, Sorenson O (2004) Science as a map in technological search. *Strategic Management J.* 25(8–9):909–928.
Fleming L, King C III, Juda AI (2007) Small worlds and regional innovation. *Organ. Sci.* 18(6):938-954.
Fleming L, Greene H, Li GC, Marx M, Yao D (2019) Government-funded research increasingly fuels innovation. *Science.* 364(6446):1139–1141.
Gittelman M, Kogut B (2003) Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *Management Sci.* 49(4):366–382.
Granovetter M (1973) The Strength of Weak Ties. *Amer. J. Sociol.* 78(6):1360-1380.
Hansen MT (1999) The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Admin. Sci. Quart.* 44(1):82–111.
Hengel E (2022) Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review. *Econom. J.* 132(648):2951–2991.
Henrekson M, Rosenberg N (2001) Designing efficient institutions for science-based entrepreneurship: Lesson from the US and Sweden. *J. Technol. Transfer* 26(3):207-231.
Jensen R, Thursby M (2001) Proofs and Prototypes for Sale: The Licensing of University Inventions. *Amer. Econom. Rev.* 91(1):240–259.
King G, Tomz M, Wittenberg J (2000) Making the most of statistical analyses: Improving interpretation and presentation. *Amer. J. Polit. Sci.* 44(2):341-355.
Lane P, Lubatkin M (1998) Relative absorptive capacity and interorganizational learning. *Strategic Management J.* 19(5):461-477.
Lee JJ (2010) Heterogeneity, brokerage, and innovative performance: Endogenous formation of collaborative inventor networks. *Organ. Sci.* 21(4):804-822.
Liebeskind JP, Oliver AL, Zucker L, Brewer M (1996) Social networks, learning, and flexibility: Sourcing scientific knowledge in new biotechnology firms. *Organ. Sci.* 7(4):428-443.
Lissoni F (2010) Academic inventors as brokers. *Res. Policy 39*(7):843-857.
Marx M, Fuegi A (2020) Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management J.* 41(9):1572–1594.
Marx M, Hsu DH (2022) Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from 'Twin' Discoveries. *Management Sci.* 68(2):1330–1352.
Marx M (2026) Does Patenting Promote the Progress of Science? Evidence from a public database of patent paper pairs. Working paper, Cornell University, Ithaca, NY.
Merton RK (1957) Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *Amer. Sociol. Rev.* 22(6):635–659.
Merton RK (1961) Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science. *Proc. Amer. Philos. Soc.* 105(5):470–486.

Mohnen M (2022). Stars and brokers: Knowledge spillovers among medical scientists. *Management Sci. 68*(4):2513-2532.
Mokyr J (2002) *The Gifts of Athena: Historical Origins of the Knowledge Economy* (Princeton: Princeton University Press).
Moreira S, Markus A, Laursen K (2018) Knowledge diversity and coordination: The effect of intrafirm inventor task networks on absorption speed. *Strategic Management J. 39*(9):2517-2546.
Mumtaz S (2025) Lost in Translation? Science Communication and the Commercial Diffusion of Ideas. Working paper, UC Berkeley, Berkeley, CA.
Murray F (2004) The role of academic inventors in entrepreneurial firms: Sharing in the laboratory life. *Res. Policy 33*(4):643-659.
National Science Board, National Science Foundation (2023) Publications output: U.S. trends and international comparisons. Report, Science and Engineering Indicators 2024, NSB-2023-33, Alexandria, VA.
Nelson RR (1959) The Simple Economics of Basic Scientific Research. *J. Political Econom.* 67(3):297–306.
Operti E, Carnabuci G (2014) Public knowledge, private gain: The effect of spillover networks on firms' innovative performance. *J. Management* 40(4):1042-1074.
Owen-Smith J, Powell W (2004) Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organ. Sci.* 15(1):5-21.
Paruchuri S, Awate S (2017) Organizational knowledge networks and local search: The role of intra-organizational inventor networks. *Strategic Management J. 38*(3):657-675.
Pavitt K (1991) What makes basic research economically useful? *Res. Policy 20*(2):109-119.
Perkmann M, Salandra R, Tartari V, McKelvey M, Hughes A (2021) Academic engagement: A review of the literature 2011-2019. *Res. Policy 50*(1):104114.
Perkmann M, Tartari V, McKelvey M, Autio E, Broström A, D'este P, Fini R, Geuna A, Grimaldi R, Hughes A, Krabel S, Kitson M, Llerena P, Lissoni F, Salter A, Sobrero M (2013) Academic engagement and commercialisation: A review of the literature on university–industry relations. *Res. Policy 42*(2):423-442.
Roach M, Cohen WM (2013) Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research. *Management Sci.* 59(2):504-525.
Roche MP (2023) Academic entrepreneurship: Entrepreneurial advisors and their advisees' outcomes. *Organ. Sci.* 34(2):959-986.
Rosenberg N (1990) Why do firms do basic research (with their own money)? *Res. Policy* 19(2):165-174.
Sauermann H, Roach M (2012) Science PhD career preferences: Levels, changes, and advisor encouragement. *PLoS ONE* 7(5):1–9.
Schaper T, Arts S, Veugelers R (2025) Not like the others: Frontier scientists for inventive performance. *Res. Policy* 54(10):105339.
Scharfmann E, Marx M, Fleming L (2025) Pasteur's quadrant researchers bring novelty, impact to publishing, and patenting. *Science* 390(6776):891–893.
Schoenberger S, Toes R, van der Voort E, Offringa R, Melief C (1998) T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions. *Nature* 393(6684):480–483.
Shane S (2004) *Academic entrepreneurship: University spinoffs and wealth creation* (Cheltenham: Edward Elgar Publishing).
Simeth M, Cincera M (2016) Corporate science, innovation, and firm value. *Management Sci.* 62(7):1970-1981.
Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Sci.* 51(5):756–770.
Smalheiser NR, Torvik VI (2009). Author name disambiguation. *Annu. Rev. Inform. Sci. Technol.* 43(1):1–43.
Torvik VI (2018) Author-Linked data for Author-ity 2009. Dataset, University of Illinois Urbana-Champaign, Champaign, IL.

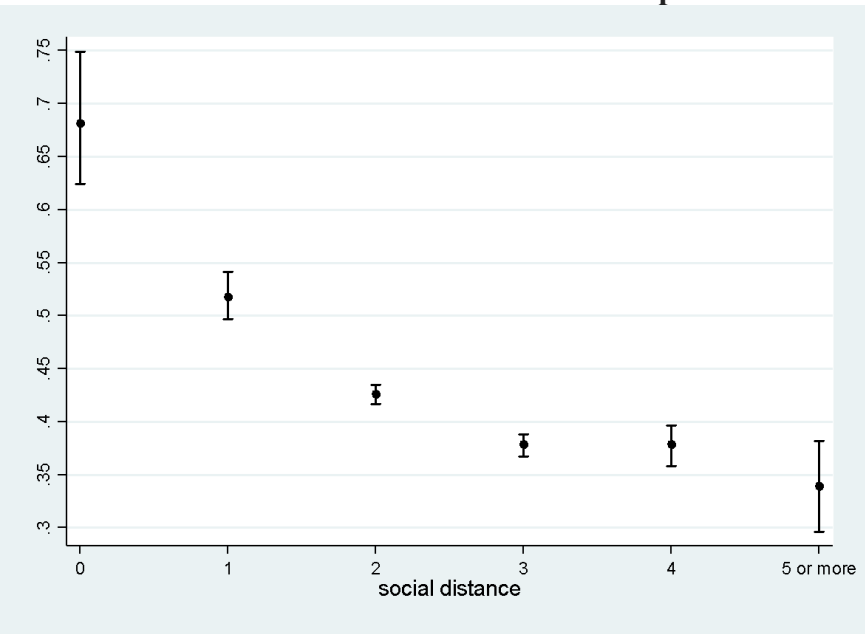
Torvik VI (2021) MapAffil 2018 dataset -PubMed author affiliations mapped to cities and their geocodes worldwide with extracted disciplines, inferred GRIDs, and assigned ORCIDs. Dataset, University of Illinois Urbana-Champaign, Champaign, IL.
Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM Trans. Knowl. Discov. Data* 3(3):1–30.
Torvik VI, Smalheiser NR (2018) Author-ity 2009—PubMed author name disambiguated dataset Dataset, University of Illinois Urbana-Champaign, Champaign, IL.
Tushman M (1977) Special Boundary Roles in the Innovation Process. *Admin. Sci. Quart.* 22(4):587-605.
Tushman ML, Scanlan TJ (1981) Boundary spanning individuals: Their role in information transfer and their antecedents. *Acad. Management J. 24*(2):289-305.
Vakili K, Teodoridis F, Bikard M (2022) Detrimental collaborations in creative work: Evidence from economics. *Organ. Sci.* 33(5):1741-1755.
Wasserman S, Faust K (1994) *Social Network Analysis* (New York: Cambridge University Press).
Zucker LG, Darby MR, Brewer MB (1998) Intellectual human capital and the birth of US biotechnology enterprises. *Amer. Econom. Rev.* 88(1):290-306.
Zucker LG, Darby MR, Armstrong JS (2002) Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology. *Management Sci.* 48(1):138–153.

Figure 1. Shortest interpersonal paths between corporate inventors on a citing patent and academic scientists on cited and uncited twin papers



Notes: The graph depicts all direct collaborators of the corporate inventors of patent US7288251 (titled "Antibodies to CD40") assigned to Abgenix and Pfizer, as well as all direct collaborators of the academic authors of the cited paper by Bennett et al. (1998) titled "Help for cytotoxic-T-cell responses is mediated by CD40 signalling," and all direct collaborators of the academic authors of the uncited twin paper by Schoenberger et al. (1998) titled "T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions." Each node represents an individual: red triangles indicate individuals who exclusively publish, yellow dots indicate individuals who only patent, and green squares indicate individuals who both publish and patent. Corporate inventors of the citing patent are represented by black light bulbs, academic authors of the cited paper are shown as checked drawing boards, and authors of the uncited twin paper are depicted as unchecked drawing boards. The edges (lines connecting two nodes) represent co-authorship/co-inventorship ties up to 2002, the patent filing year. To improve visibility, ties established before 1989 are removed. The length of the ties is arbitrary. The shortest path between the corporate inventors of the citing patent and the academic scientists of the cited paper has a length of 1 and is highlighted as a bold line. The shortest path between the corporate inventors of the citing patent and the academic scientists of the uncited twin paper has a length of 2 but is not displayed.

Figure 2. Predicted probability of corporate patent citations to academic papers across different levels of social distance between academic authors and corporate inventors



Notes: Figure 2 depicts the predicted probability of citation, along with 95% confidence intervals, across different levels of social distance between academic authors and corporate inventors in paper-patent dyads, holding all other covariates at their sample means. The predicted probabilities and corresponding confidence intervals are obtained using the methodology proposed by King et al. (2000). Detailed regression results are presented in column (4-2) of Table 4.

Table 1. Main variables and definitions

Variable	Definition
cited	Dummy equal to 1 if paper is cited by patent.
social_dist	Shortest path between any of the author(s) of the paper and any of the inventor(s) of the patent at the patent filing year *t*. Unreachable is imputed with value 9, i.e., one more than the maximum distance observed in our data, i.e., distance 8.
social_dist_0	Dummy equal to 1 if shortest path is 0, i.e., an author of the paper is inventor of the patent.
social_dist_1	Dummy equal to 1 if shortest path is 1, i.e., an author of the paper is a past collaborator of an inventor of the patent.
social_dist_2	Dummy equal to 1 if shortest path is 2, an author of the paper and an inventor of the patent have a common collaborator.
social_dist_3	Dummy equal to 1 if shortest path is 3.
social_dist_4	Dummy equal to 1 if shortest path is 4.
social_dist_5_or_more	Dummy equal to 1 if shortest path is 5 or more, including unreachable.
soc_distance_time	Weighted shortest path length (P^*) between any author of paper *a* and any inventor of patent *b*. The path consists of 1 to k connections, where each connection (edge) is weighted based on the time since the most recent collaboration between the individuals in the path. The weight is 1 plus the difference between the patent filing year (*t*) and the last collaboration year. If no path exists, the length is set to 30, one more than the maximum observed distance (29). If an author of the paper is also an inventor of the patent, the distance is 0. The shortest path is computed using Dijkstra's algorithm from the R igraph library. $$soc_distance_time_{ab} = min_{P^* \in P(a,b)} \sum_{i=1}^{k} (weight(a_{i-1}, a_i) + 1)$$
degree1_size	Number of collaborators (those within 1 degree of separation) of the paper's authors up to the year of patent filing (year *t*).
degree2_size	Number of connections within 2 degree of separation of the paper's authors up to the year of patent filing (year *t*).
degree3_size	Number of connections within 3 degree of separation of the paper's authors up to the year of patent filing (year *t*).
no. authors	Number of authors on the paper.
author-inventor team	Dummy equal to 1 if at least one of the paper's authors is an author-inventor at the patent filing year *t*.
publication stock	Number of unique PubMed papers published by any of the paper's authors up to the patent filing year *t*
patent stock	Number of unique granted patents by any of the paper's authors up to the patent filing year *t*.
institutional prestige	Count of unique PubMed papers linked to individuals with the same affiliations as the authors on the focal paper up to the year of patent filing (year *t*).
U.S. based	Dummy equal to 1 if any affiliation of the authors on the focal paper is in the US.
min geographic distance	Minimum geographic distance in miles between any of the author(s) of the paper and any of the inventor(s) of the patent.
journal impact factor	Journal impact factor of the paper in year of publication
journal commercial impact factor	Journal commercial impact factor of the paper in year of publication based on patent citations to papers from the same journal as the focal paper (see Bikard and Marx (2020)).
paper citations	Number of times the focal paper is cited by other papers up to the year before patent filing (t-1).
patent citations	Number of times the focal paper is cited by patents up to the year before patent filing (t-1).
time lag	Number of days between publication date of paper and filing date of patent.
published first	Dummy equal to 1 if the focal paper is published earlier or on the same day as the twin paper(s).
text length	Number of words in title and abstract of paper.
readability	The Gunning Fog index, a measure of readability, is calculated by considering the average number of words per sentence and the proportion of complex words (defined as words with three or more syllables) This calculation uses both the title and abstract of a paper, applying the formula: 0.4×(average sentence length+100×(complex words/total words)) (Hengel 2022).
active_inventor_author	Dummy equal to 1 if at least one of the patent's inventors published a paper in the year before the patent filing year *t*.
close_paper	We compute the SPECTER embedding of each paper using its title and abstract, and measure the cosine similarity between each prior paper of the corporate inventors on the patent published before the patent filing year and the twin papers (Cohan et al. 2020). Next, we calculate the maximum of these pairwise paper similarities at the twin-patent level as a measure of how close and familiar the corporate inventors are to the academic twin discovery. We then create a binary indicator, close_paper, which equals one when this value exceeds the median in our dataset.

Notes: The unit of observation is a paper-patent dyad, which includes both cited paper-patent dyads and uncited paper-patent dyads. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. We do not apply the logarithmic transformation to *social_dist* to maintain ease of interpretation.

Table 2. Summary statistics at the paper level

	All		Cited		Not Cited		T test
	Mean	SD	Mean	SD	Mean	SD	p value
degree1_size	5.532	1.070	5.579	1.100	5.486	1.039	0.289
degree2_size	9.448	1.286	9.501	1.320	9.395	1.252	0.321
degree3_size	12.822	1.129	12.861	1.152	12.784	1.106	0.411
no. authors	1.535	0.590	1.571	0.569	1.499	0.609	0.143
author-inventor team	0.642	0.480	0.670	0.471	0.614	0.488	0.153
publication stock	5.011	0.990	5.052	1.025	4.970	0.952	0.314
patent stock	1.109	1.091	1.202	1.119	1.017	1.056	0.039**
institutional prestige	9.535	1.438	9.593	1.395	9.478	1.480	0.334
U.S. based	0.637	0.481	0.646	0.479	0.627	0.484	0.630
journal impact factor	1.817	0.677	1.833	0.679	1.802	0.677	0.577
journal commercial impact factor	0.110	0.113	0.114	0.118	0.107	0.108	0.448
paper citations	3.175	1.893	3.221	1.877	3.129	1.912	0.558
patent citations	0.625	0.864	0.615	0.828	0.635	0.900	0.786
published first	0.699	0.459	0.714	0.453	0.685	0.465	0.435
text length	5.169	0.420	5.171	0.421	5.168	0.421	0.923
readability	2.740	0.184	2.738	0.177	2.742	0.191	0.784

Notes: The unit of observation is a paper (n=501) from 247 paper twins (7 are triplets), including cited papers (n=294) and uncited twin papers (n=295). The large majority of the papers (413, or 82%) maintain a consistent status as either cited or uncited throughout the dataset. However, 88 papers (18%) appear both as cited and uncited paper due to multiple patent citations to different papers of the same twin. Excluding these dual-role papers from the analysis does not significantly alter the summary statistics or T-test results. We consider the first observation year of each paper in our dataset, whether cited or uncited. T-tests assume unequal variances. *, **, and *** denote significance at the 10%, 5%, and 1% level. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values.

Table 3. Summary statistics at the paper-patent dyad level

	All		Cited		Not Cited		T test
	Mean	SD	Mean	SD	Mean	SD	p value
cited	0.497	0.500	1.000	0.000	0.000	0.000	
social_dist	2.789	1.145	2.751	1.195	2.827	1.091	0.034**
social_dist_0	0.016	0.124	0.028	0.164	0.004	0.062	0.000***
social_dist_1	0.053	0.224	0.063	0.243	0.043	0.204	0.005***
social_dist_2	0.271	0.445	0.257	0.437	0.285	0.452	0.036**
social_dist_3	0.538	0.499	0.527	0.499	0.548	0.498	0.168
social_dist_4	0.100	0.300	0.104	0.305	0.096	0.294	0.383
social_dist_5_or_more	0.023	0.149	0.022	0.147	0.023	0.151	0.808
soc_distance_time	5.912	3.848	5.979	3.811	5.847	3.884	0.269
min geographic distance	6.416	2.328	7.027	1.938	5.813	2.516	0.000***
time lag	8.063	1.169	8.077	1.122	8.050	1.213	0.451

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. T-tests assume unequal variances. *, **, and *** denote significance at the 10%, 5%, and 1% level. Min geographic distance and time lag are logarithmically transformed after adding one.

Table 4. Influence of social distance between academic scientists and corporate inventors on company patent citation to academic paper: Binary indicators for social distance

	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Linear probability
	Only controls		Exclude social distance 0	Exclude social distance 0 or 1	Exclude social unreachable	Social distance measured before academic paper twin	Exclude corporate funding	Exclude front page citations	
	(4-1)	(4-2)	(4-3)	(4-4)	(4-5)	(4-6)	(4-7)	(4-8)	(4-9)
social_dist_1		-2.571***			-2.588***	-2.822***	-2.644***	-2.744***	-0.348***
		(0.653)			(0.649)	(0.622)	(0.769)	(0.768)	(0.088)
social_dist_2		-4.101***	-1.775***		-4.121***	-2.756***	-4.206***	-4.386***	-0.660***
		(0.974)	(0.495)		(0.968)	(0.737)	(1.189)	(1.123)	(0.138)
social_dist_3		-4.910***	-2.625***	-0.852***	-4.914***	-3.631***	-5.040***	-5.294***	-0.836***
		(1.068)	(0.631)	(0.254)	(1.059)	(0.785)	(1.316)	(1.265)	(0.158)
social_dist_4		-4.906***	-2.706***	-0.966**	-4.893***	-4.295***	-5.222***	-5.257***	-0.812***
		(1.149)	(0.722)	(0.388)	(1.139)	(1.021)	(1.375)	(1.434)	(0.195)
social_dist_5_or_more		-5.585***	-3.448***	-1.759***	-5.522***	-5.098***	-5.876***	-5.127***	-1.096***
		(1.111)	(0.745)	(0.607)	(1.113)	(1.095)	(1.338)	(1.503)	(0.200)
N of observations (paper-patent dyads)	4,169	4,169	4,039	3,706	4,099	4,169	3,591	3,422	5,745
N of twin-patent dyads	2,068	2,068	2,003	1,837	2,033	2,068	1,782	1,698	2,850
N of citing patents	1,951	1,951	1,886	1,732	1,920	1,951	1,704	1,603	2,641
N of papers	501	501	493	462	491	501	405	362	634
N of twins	247	247	243	228	242	247	200	178	312
Log-Likelihood	-945.8	-847.4	-821.5	-724.6	-834.0	-874.3	-713.4	-666.5	
R2	0.346	0.414	0.414	0.437	0.414	0.396	0.428	0.439	0.393
Mean(cited)	0.497	0.497	0.496	0.496	0.496	0.497	0.497	0.497	0.635

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports log-odds coefficient estimates from conditional logit models with fixed effects at the twin patent-level, social_dist_0 is the excluded category in columns 2, 5, 6, 7, 8, 9 and *social_dist_1* is the excluded category in column 3, *social_dist_2* is the excluded category in column 4. Column (4-9) also includes cases where a patent cites all twin papers, estimated using a linear probability model with twin patent-level fixed effects. This extended sample includes 312 paper twins, 634 papers, and 2,641 citing patents. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table 5. Influence of social distance between academic scientists and corporate inventors on company patent citation to academic paper: Count measure for social distance

	Conditional logit	Conditional logit	Conditional logit Exclude social distance 0	Conditional logit Exclude social distance 0 or 1	Conditional logit Exclude social unreachable	Conditional logit Social distance measured before academic paper twin	Conditional logit Exclude corporate funding	Conditional logit Exclude front page citations	Linear probability	Conditional logit Social distance time (shortest path weighted by recency of ties)
	(5-1)	(5-2)	(5-3)	(5-4)	(5-5)	(5-6)	(5-7)	(5-8)	(5-9)	(5-10)
social_dist		-0.957***	-0.897***	-0.632***	-0.947***	-0.815***	-1.020***	-1.021***	-0.196***	-0.214***
		(0.205)	(0.208)	(0.190)	(0.204)	(0.207)	(0.226)	(0.246)	(0.041)	(0.062)
N of observations (paper-patent dyads)	4,169	4,169	4,039	3,706	4,099	4,169	3,591	3,422	5,745	4,169
N of twin-patent dyads	2,068	2,068	2,003	1,837	2,033	2,068	1,782	1,698	2,850	2,068
N of citing patents	1,951	1,951	1,886	1,732	1,920	1,951	1,704	1,603	2,641	1,951
N of papers	501	501	493	462	491	501	405	362	634	501
N of twins	247	247	243	228	242	247	200	178	312	247
Log-Likelihood	-945.8	-873.8	-839.1	-729.7	-861.1	-887.7	-731.3	-690.7		-927.4
R2	0.346	0.396	0.401	0.433	0.395	0.386	0.413	0.418	0.384	0.359
Mean(cited)	0.497	0.497	0.496	0.496	0.496	0.497	0.497	0.497	0.635	0.497

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units (except in column (5-9)). Column (5-9) also includes cases where a patent cites all twin papers, estimated using a linear probability model with twin patent-level fixed effects. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table 6. Mechanisms: Company inventors active in scientific research as boundary-spanners

	Conditional logit (6-1)	Conditional logit (6-2)	Conditional logit (6-3)	Conditional logit (6-4)
social_dist	-0.957***	-0.068	-0.376***	0.105
	(0.205)	(0.187)	(0.140)	(0.213)
social_dist*active_inventor_author		-1.081***		-0.744***
		(0.289)		(0.216)
social_dist*close_paper			-0.920***	-0.728**
			(0.325)	(0.321)
N of observations (paper-patent dyads)	4,169	4,169	4,115	4,115
N of twin-patent dyads	2,068	2,068	2,041	2,041
N of citing patents	1,951	1,951	1,924	1,924
N of papers	501	501	493	493
N of twins	247	247	243	243
Log-Likelihood	-873.8	-859.0	-848.4	-842.1
R2	0.396	0.406	0.406	0.410
Mean(cited)	0.497	0.497	0.496	0.496

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. *Active_inventor_author* is a dummy equal to 1 if any of patent's inventors have at least 1 PubMed paper published in the year before patent filing. *Close_paper* is a dummy equal to 1, if the max similarity between inventors' prior papers (published before the patent filing year) and the twin papers (cited or not cited) is higher than sample median. Paper-patent dyads where inventors have 0 prior papers have a value of 0 for cosine similarity. Paper-patent dyads for which cosine similarity could not be calculated due to missing text data are dropped in models 6-3, 6-4. We tested alternative models where patent-twin dyads with 0 prior papers were also put to missing observations, the effect size and significance hold. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

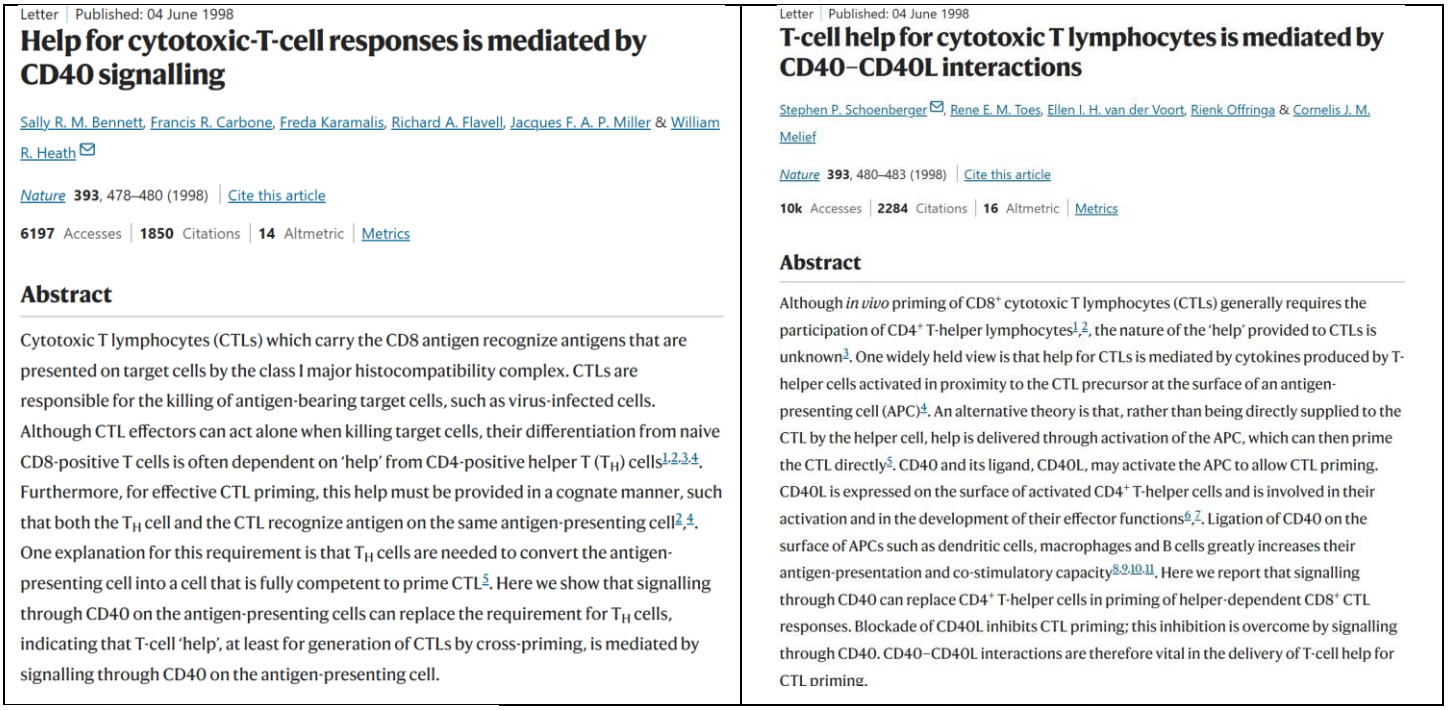
Table 7. Robustness checks

	Exclude higher order twins	Exclude twins published > 3 months apart	Exclude twins text similarity <0.90	Twins with Pubmed rank order=1	Twins with back-to-back papers	Twins with share of adjacent citations =100%	Exclude twins with papers from large teams (>5)	Exclude twins with top cited papers	Exclude patents of Genentech & Millennium Pharmaceuticals
	(7-1)	(7-2)	(7-3)	(7-4)	(7-5)	(7-6)	(7-7)	(7-8)	(7-9)
social_dist	-0.976***	-1.066***	-1.132***	-1.627***	-1.260***	-0.893***	-1.310***	-0.983***	-0.873***
	(0.209)	(0.234)	(0.217)	(0.155)	(0.282)	(0.199)	(0.193)	(0.210)	(0.185)
N of observations (paper-patent dyads)	3,996	3,368	3,310	2,010	2,301	935	2,953	4,007	3,417
N of twin-patent dyads	1,998	1,668	1,642	1,005	1,150	467	1,464	1,987	1,692
N of citing patents	1,888	1,583	1,587	980	1,114	450	1,410	1,870	1,584
N of papers	480	354	308	130	181	103	246	493	493
N of twins	240	174	152	65	90	51	121	243	243
Log-Likelihood	-836.6	-678.2	-663.2	-333.1	-431.2	-196.7	-537.9	-821.1	-803.8
R2	0.396	0.420	0.423	0.522	0.459	0.393	0.475	0.410	0.322
Mean(cited)	0.500	0.496	0.496	0.500	0.500	0.501	0.496	0.496	0.496

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Online Appendix

Figure A.1 Example of twin papers



Letter | Published: 04 June 1998

Help for cytotoxic-T-cell responses is mediated by CD40 signalling

Sally R. M. Bennett, Francis R. Carbone, Freda Karamalis, Richard A. Flavell, Jacques F. A. P. Miller & William R. Heath

Nature **393**, 478–480 (1998) | Cite this article

6197 Accesses | **1850** Citations | **14** Altmetric | Metrics

Abstract

Cytotoxic T lymphocytes (CTLs) which carry the CD8 antigen recognize antigens that are presented on target cells by the class I major histocompatibility complex. CTLs are responsible for the killing of antigen-bearing target cells, such as virus-infected cells. Although CTL effectors can act alone when killing target cells, their differentiation from naive CD8-positive T cells is often dependent on 'help' from CD4-positive helper T (T_H) cells[1,2,3,4]. Furthermore, for effective CTL priming, this help must be provided in a cognate manner, such that both the T_H cell and the CTL recognize antigen on the same antigen-presenting cell[2,4]. One explanation for this requirement is that T_H cells are needed to convert the antigen-presenting cell into a cell that is fully competent to prime CTL[5]. Here we show that signalling through CD40 on the antigen-presenting cells can replace the requirement for T_H cells, indicating that T-cell 'help', at least for generation of CTLs by cross-priming, is mediated by signalling through CD40 on the antigen-presenting cell.

Letter | Published: 04 June 1998

T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions

Stephen P. Schoenberger, Rene E. M. Toes, Ellen I. H. van der Voort, Rienk Offringa & Cornelis J. M. Melief

Nature **393**, 480–483 (1998) | Cite this article

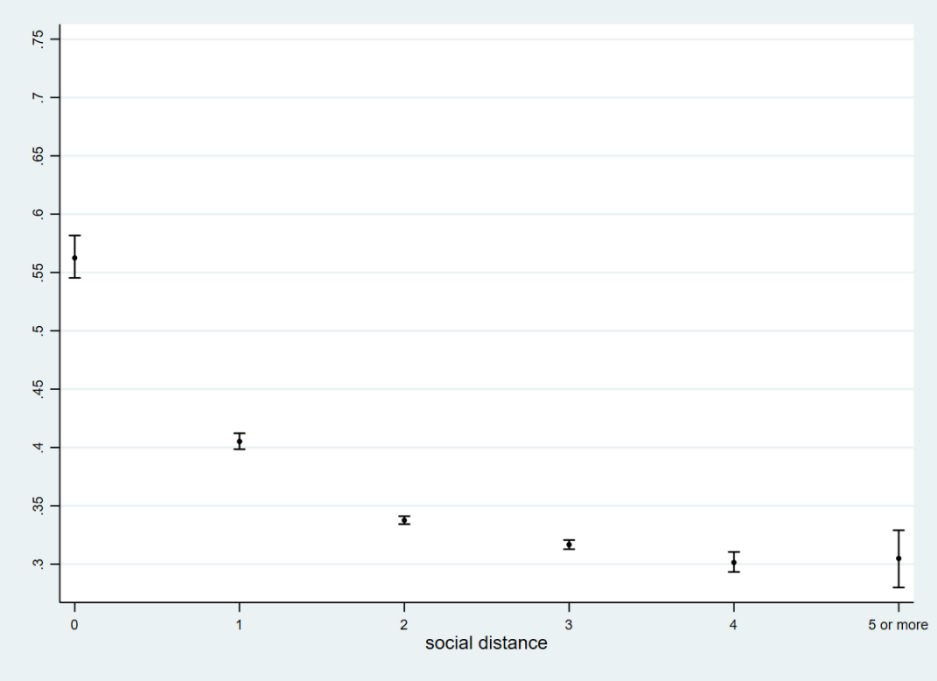
10k Accesses | **2284** Citations | **16** Altmetric | Metrics

Abstract

Although *in vivo* priming of $CD8^+$ cytotoxic T lymphocytes (CTLs) generally requires the participation of $CD4^+$ T-helper lymphocytes[1,2], the nature of the 'help' provided to CTLs is unknown[3]. One widely held view is that help for CTLs is mediated by cytokines produced by T-helper cells activated in proximity to the CTL precursor at the surface of an antigen-presenting cell (APC)[4]. An alternative theory is that, rather than being directly supplied to the CTL by the helper cell, help is delivered through activation of the APC, which can then prime the CTL directly[5]. CD40 and its ligand, CD40L, may activate the APC to allow CTL priming. CD40L is expressed on the surface of activated $CD4^+$ T-helper cells and is involved in their activation and in the development of their effector functions[6,7]. Ligation of CD40 on the surface of APCs such as dendritic cells, macrophages and B cells greatly increases their antigen-presentation and co-stimulatory capacity[8,9,10,11]. Here we report that signalling through CD40 can replace $CD4^+$ T-helper cells in priming of helper-dependent $CD8^+$ CTL responses. Blockade of CD40L inhibits CTL priming; this inhibition is overcome by signalling through CD40. CD40–CD40L interactions are therefore vital in the delivery of T-cell help for CTL priming.

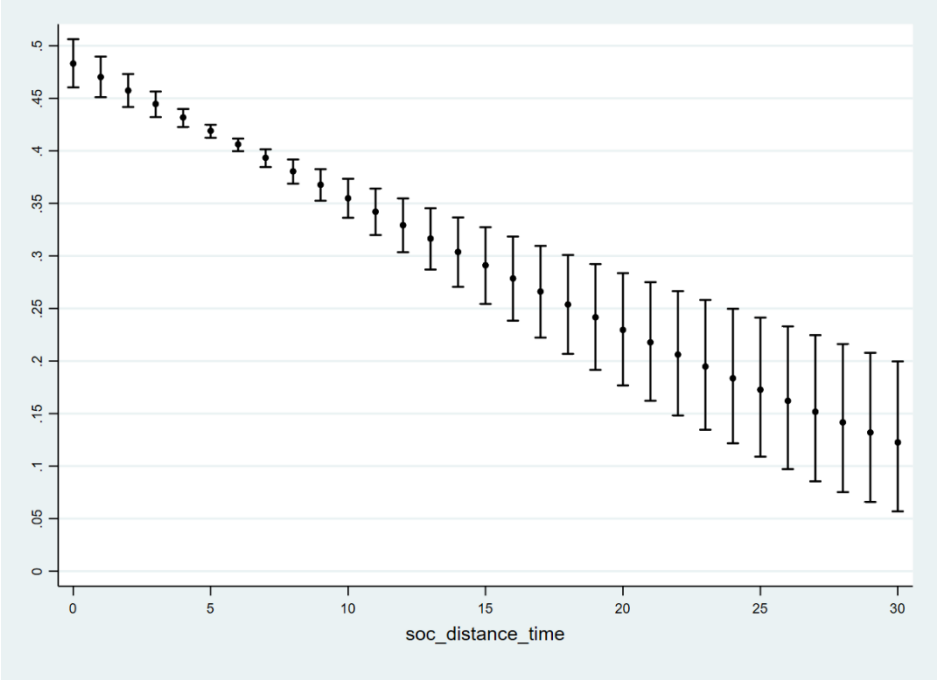
Notes: Example of twin papers published back-to-back in *Nature*. Paper on the left: Bennett S, Carbone F, Karamalis F, Flavell R, Miller J, Heath W (1998) Help for cytotoxic-T-cell responses is mediated by CD40 signalling. Nature 393(6684):478–480. Paper on the right: Schoenberger S, Toes R, van der Voort E, Offringa R, Melief,C (1998) T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions. Nature 393(6684):480–483.

Figure A.2 Predicted probability of paper-to-paper citations across different levels of social distance between citing and cited academic authors



Notes: Figure A.2 depicts the predicted probability of citation, along with 95% confidence intervals, across different levels of social distance between authors in paper-to-paper dyads. The predicted probabilities and corresponding confidence intervals are obtained using the methodology proposed by King et al. (2000). Detailed regression results are available from the authors.

Figure A.3 Predicted probability of corporate patent citations to academic papers across different levels of *soc_distance_time* between academic authors and corporate inventors



Notes: Figure A.3 depicts the predicted probability of citation along with the 95% confidence interval across different levels of soc distance time (shortest path weighted by recency of ties) between academic authors and corporate inventors in paper-patent dyads. The predicted probabilities and corresponding confidence intervals are obtained using the methodology proposed by King et al. (2000). Detailed regression results are presented in column (5-10) of Table 5.

Table A.1 Universities and academic institutions most commonly listed as affiliation on twin papers

	N. (1)	Share (2)
Harvard University	35	0.070
Massachusetts Institute of Technology	15	0.030
University of California, San Diego	15	0.030
University of California, San Francisco	13	0.026
Yale University	13	0.026
Johns Hopkins University	11	0.022
National Cancer Institute	11	0.022
University of Toronto	10	0.020
Cold Spring Harbor Laboratory	10	0.020
Massachusetts General Hospital	10	0.020
University of Washington	10	0.020
CNRS	10	0.020
University of California at Berkeley	10	0.020
Stanford University	9	0.018
European Molecular Biology Laboratory	9	0.018
Dana Farber Cancer Institute	9	0.018
University of Chicago	8	0.016
University of Michigan	8	0.016
Baylor College of Medicine	8	0.016
California Institute of Technology	7	0.014

Notes: Top 20 universities and institutions most commonly listed as affiliation on twin papers. The number in column 1 refers to the number of papers and the share in column 2 is computed based on the total number of twin papers in our sample (n=501). Papers with multiple affiliations can appear in multiple rows.

Table A.2 Companies most commonly listed as assignee on citing patents

	N. (1)	Share (2)
Genentech	219	0.112
Millennium Pharmaceuticals	209	0.107
ZymoGenetics	59	0.030
BASF Plant Science	45	0.023
BASF AG	35	0.018
CuraGen Corporation	34	0.017
Eli Lilly & Company	32	0.016
Kyowa Hakko Kogyo Co.	26	0.013
Amgen	25	0.013
Kyowa Hakko Kirin Co.	24	0.012
Schering Corporation	23	0.012
Virogenetics Corporation	23	0.012
Abbott Laboratories	22	0.011
Agrinomics	22	0.011
Biogen IDEC	20	0.010
Human Genome Sciences	19	0.010
Bio Merieux	17	0.009
Bristol-Myers Squibb Company	18	0.009
Pioneer Hi-Bred International	16	0.008
Geron Corporation	15	0.008

Notes: Top 20 companies most commonly listed as assignee on citing patents. The number in column 1 refers to the number of patents and the share in column 2 is computed based on the total number of citing patents in our sample (n=1,951). Patents with multiple assignees can appear in multiple rows.

Table A.3 Results by company type

	Main effect	U.S. public firms	Firm age <=3 years	Firm age <=5 years	Firm age <=10 years	Firm patent count<p50	Firm patent count>=p50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
social_dist	-0.957***	-0.813***	-0.952*	-0.656*	-0.661***	-0.948***	-1.011***
	(0.205)	(0.151)	(0.526)	(0.342)	(0.221)	(0.320)	(0.225)
N of observations (paper-patent dyads)	4,169	2,124	237	405	841	765	1,288
N of twin-patent dyads	2,068	1,050	115	198	411	370	644
N of citing patents	1,951	1,011	103	181	387	353	626
N of papers	501	245	92	130	202	175	142
N of twins	247	121	44	63	99	86	71
Log-Likelihood	-873.8	-366.0	-45.5	-72.3	-162.2	-147.4	-155.1
R2	0.396	0.504	0.449	0.487	0.446	0.447	0.653
Mean(cited)	0.497	0.494	0.489	0.491	0.490	0.485	0.500

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table A.4 Robustness checks: Company inventors active in scientific research as cross-institutional boundary-spanners

	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Linear probability	Conditional logit	Conditional logit
			Exclude social distance 0	Exclude social distance 0 or 1	Exclude social unreachable	Social distance measured before academic twin	Exclude corporate funding	Exclude front page citations		Social distance time (shortest path weighted by recency of ties)	Count measure for *active_inventor_author*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
social_dist		-0.068	-0.061	-0.001	-0.067	-0.189	-0.162	0.003	-0.009	-0.027	-0.764***
		(0.187)	(0.195)	(0.229)	(0.185)	(0.253)	(0.165)	(0.233)	(0.040)	(0.098)	(0.173)
social_dist*active_inventor_author		-1.081***	-1.032***	-0.849**	-1.072***	-0.718**	-1.018***	-1.185***	-0.223***	-0.218	-0.034**
		(0.289)	(0.299)	(0.356)	(0.287)	(0.295)	(0.318)	(0.326)	(0.056)	(0.134)	(0.014)
N of observations (paper-patent dyads)	4,169	4,169	4,039	3,706	4,099	4,169	3,591	3,422	5,745	4,169	4,169
N of twin-patent dyads	2,068	2,068	2,003	1,837	2,033	2,068	1,782	1,698	2,850	2,068	2,068
N of citing patents	1,951	1,951	1,886	1,732	1,920	1,951	1,704	1,603	2,641	1,951	1,951
N of papers	501	501	493	462	491	501	405	362	634	501	501
N of twins	247	247	243	228	242	247	200	178	312	247	247
Log-Likelihood	-945.8	-859.0	-826.9	-722.9	-846.6	-882.4	-721.1	-679.6		-924.6	-868.0
R2	0.346	0.406	0.410	0.438	0.405	0.390	0.421	0.428	0.390	0.361	0.400
Mean(cited)	0.497	0.497	0.496	0.496	0.496	0.497	0.497	0.497	0.635	0.497	0.497

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units (except in column 9). Column 9 also includes cases where a patent cites all twin papers, estimated using a linear probability model with twin patent-level fixed effects. *Active_inventor_author* is a dummy equal to 1 if any of the patent's inventors have at least 1 PubMed paper published in the year before patent filing. In column 11, *Active_inventor_author* is a total count of PubMed indexed papers authored by patent's inventors in the year before patent filing. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table A.5 Robustness checks: Company inventors with closely related scientific expertise as cross-institutional boundary-spanners

	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Linear probability	Conditional logit	Conditional logit
			Exclude social distance 0	Exclude social distance 0 or 1	Exclude social unreachable	Social distance measured before academic twin	Exclude corporate funding	Exclude front page citations		Social distance time (shortest path weighted by recency of ties)	Continuous measure for *close_paper*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
social_dist		-0.376***	-0.382***	-0.250	-0.367***	-0.206	-0.398***	-0.311*	-0.084**	0.093	0.573
		(0.140)	(0.141)	(0.187)	(0.140)	(0.207)	(0.153)	(0.164)	(0.033)	(0.069)	(0.556)
social_dist*close_paper		-0.920***	-0.915***	-0.880*	-0.922***	-0.894*	-0.935**	-1.101***	-0.165**	-0.476***	-1.948**
		(0.325)	(0.328)	(0.455)	(0.322)	(0.490)	(0.368)	(0.322)	(0.075)	(0.116)	(0.805)
N of observations (paper-patent dyads)	4,169	4,115	3,985	3,652	4,053	4,115	3,551	3,380	5,675	4,115	4,115
N of twin-patent dyads	2,068	2,041	1,976	1,810	2,010	2,041	1,762	1,677	2,815	2,041	2,041
N of citing patents	1,951	1,924	1,859	1,705	1,897	1,924	1,684	1,582	2,606	1,924	1,924
N of papers	501	493	485	454	485	493	401	356	628	493	493
N of twins	247	243	239	224	239	243	198	175	309	243	243
Log-Likelihood	-945.8	-848.4	-815.4	-712.0	-836.4	-864.5	-711.9	-666.3		-895.3	-856.7
R2	0.346	0.406	0.410	0.438	0.405	0.395	0.422	0.432	0.387	0.373	0.400
Mean(cited)	0.497	0.496	0.496	0.496	0.496	0.496	0.497	0.497	0.635	0.496	0.496

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units (except in column 9). Column 9 also includes cases where a patent cites all twin papers, estimated using a linear probability model with twin patent-level fixed effects. *Close paper* is a dummy equal to 1, if the max similarity between inventors' prior papers (published before the patent filing year) and the twin papers (cited or not cited) is higher than sample median (0.772). In column 11, *Close_paper* is a continuous measure, i.e., max cosine similarity between inventors' prior papers (published before the patent filing year) and the twin papers (cited or not cited). Paper-patent dyads where inventors have 0 prior papers have a value of 0 for cosine similarity. Paper-patent dyads for which cosine similarity could not be calculated due to missing text data are dropped in all models. We tested alternative models where patent-twin dyads with 0 prior papers were also put to missing observations, the effect size and significance hold. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table A.6 Summary statistics at the paper-paper dyad level

	All		Cited		Not Cited		T test
	Mean	SD	Mean	SD	Mean	SD	p value
cited	0.499	0.500	1.000	0.000	0.000	0.000	
social_dist	2.325	1.051	2.212	1.120	2.438	0.965	0.000***
social_dist_0	0.047	0.212	0.079	0.270	0.015	0.123	0.000***
social_dist_1	0.124	0.330	0.138	0.345	0.110	0.313	0.000***
social_dist_2	0.392	0.488	0.375	0.484	0.410	0.492	0.000***
social_dist_3	0.358	0.479	0.335	0.472	0.380	0.485	0.000***
social_dist_4	0.067	0.250	0.062	0.242	0.072	0.258	0.000***
social_dist_5_or_more	0.012	0.107	0.011	0.103	0.013	0.111	0.069*
min geographic distance	6.641	2.284	6.348	2.509	6.933	1.994	0.000***
time lag	7.417	1.110	7.433	1.053	7.401	1.164	0.002**

Notes: The unit of observation is a paper–paper dyad (n = 44,220), including cited dyads (n = 22,069) and uncited dyads (n = 22,151). The sample comprises 221 paper twins, 448 twin papers, and 21,571 citing papers. We identify all PubMed papers published between 1975 and 2009 that cite at least one of the twin papers in our sample. Following the patent-to-paper identification strategy, we retain only cases where a citing paper references one (or a subset, in the case of triplets) of the twin papers, using the uncited twin paper matched to the same citing paper as a counterfactual or "pseudo-citation." Minimum geographic distance and time lag are logarithmically transformed after adding one. T-tests assume unequal variances. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A.7 Influence of social distance on paper-to-paper citation

	Conditional logit	Conditional logit	Conditional logit Exclude social distance 0	Conditional logit Exclude social distance 0 or 1	Conditional logit Exclude social unreachable	Linear Probability
	(1)	(2)	(3)	(4)	(5)	(6)
social_dist		-0.406***	-0.298***	-0.153***	-0.407***	-0.169***
		(0.026)	(0.027)	(0.032)	(0.026)	(0.011)
N of observations (paper-paper dyads)	44,220	44,220	40,058	32,030	44,032	44,220
N of twin-citing paper dyads	22,005	22,005	19,939	15,945	21,911	22,005
N of citing papers	21,571	21,571	19,562	15,703	21,479	21,571
N of twin papers	448	448	448	448	448	448
N of twins	221	221	221	221	221	221
Log-Likelihood	-13890.6	-13620.1	-12699.6	-10241.8	-13560.6	
R2	0.094	0.112	0.086	0.078	0.112	0.125
Mean(cited)	0.499	0.499	0.499	0.499	0.499	0.499

Notes: The unit of observation is a paper–paper dyad (n = 44,220), including cited (n = 22,069) and uncited (n = 22,151) dyads. The sample includes 221 paper twins, 448 twin papers, and 21,571 citing papers. We identify all cases where a citing paper references only one (or a subset, in the case of triplets) of the twin papers. The binary outcome variable *cited* equals 1 when the cited paper is referenced by the citing paper and 0 for the uncited twin paper matched to the same citing paper. The table reports log-odds coefficient estimates from conditional logit models with fixed effects at the twin–citing-paper level (except in column 6). Column 6 is estimated using a linear probability model with twin–citing-paper fixed effects. All models include the same control variables as in the patent-to-paper analysis: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table A.8 Robustness checks: Binary indicators for social distance

	Exclude higher order twins	Exclude twins published > 3 months apart	Exclude twins text similarity <0.90	Twins with share of adjacent citations =100%	Exclude twins with papers from large teams (>5)	Exclude twins with top cited papers	Exclude patents of Genentech & Millennium Pharmaceuticals
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
social_dist_1	-2.568***	-2.494**	-3.331***	-1.742**	-3.138***	-2.532***	-2.277***
	(0.665)	(1.031)	(0.717)	(0.686)	(0.604)	(0.691)	(0.626)
social_dist_2	-4.104***	-4.089***	-5.232***	-3.510***	-5.655***	-3.985***	-3.631***
	(0.992)	(1.307)	(1.053)	(0.744)	(0.866)	(1.053)	(0.853)
social_dist_3	-4.953***	-5.152***	-6.152***	-4.234***	-6.687***	-4.884***	-4.423***
	(1.075)	(1.433)	(1.188)	(0.668)	(0.981)	(1.130)	(0.895)
social_dist_4	-4.918***	-5.309***	-6.109***	-3.701***	-6.861***	-4.944***	-4.395***
	(1.182)	(1.515)	(1.327)	(0.839)	(1.095)	(1.204)	(0.963)
social_dist_5_or_more	-5.741***	-6.160***	-6.836***	9.905***	-7.183***	-5.676***	-5.042***
	(1.134)	(1.293)	(1.153)	(1.823)	(1.102)	(1.145)	(1.058)
N of observations (paper-patent dyads)	3,996	3,368	3,310	935	2,953	4,007	3,417
N of twin-patent dyads	1,998	1,668	1,642	467	1,464	1,987	1,692
N of citing patents	1,888	1,583	1,587	450	1,410	1,870	1,584
N of papers	480	354	308	103	246	493	493
N of twins	240	174	152	51	121	243	243
Log-Likelihood	-810.2	-663.5	-632.8	-187.5	-512.4	-799.5	-783.7
R2	0.415	0.433	0.449	0.422	0.500	0.425	0.339
Mean(cited)	0.500	0.496	0.496	0.501	0.496	0.496	0.496

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Coefficients are in log-odds units. Distance 0 is excluded category in all models. All models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. Please notice that estimations for the subsample of twins with PubMed rank order=1 or published back-to-back did not converge because of insufficient observations for some of the different binary categories of social distance. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level. Mean(cited) is the observed citation probability in each column's estimation sample.

Table A.9 Binary social-distance estimates with and without the geographic proximity control

	Conditional logit Include min geographic distance (1)	Conditional logit Exclude min geographic distance (2)
social_dist_1	-2.571***	-1.357**
	(0.653)	(0.631)
social_dist_2	-4.101***	-2.865***
	(0.974)	(0.803)
social_dist_3	-4.910***	-3.488***
	(1.068)	(0.801)
social_dist_4	-4.906***	-3.468***
	(1.149)	(0.885)
social_dist_5_or_more	-5.585***	-4.492***
	(1.111)	(0.997)
N of observations (paper-patent dyads)	4,169	4,169
N of twin-patent dyads	2,068	2,068
N of citing patents	1,951	1,951
N of papers	501	501
N of twins	247	247
Log-Likelihood	-847.4	-895.4
R2	0.414	0.381

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level, social_dist_0 is the excluded category in columns 1 and 2. Both models include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. First model additionally includes *min geographic distance* as a control variable. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A.10 Count social-distance estimates with and without the geographic proximity control

	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit	Conditional logit
	Include min geographic distance				Exclude min geographic distance			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
social_dist	-0.957***	-0.068	-0.376***	0.105	-0.775***	0.010	-0.293*	0.173
	(0.205)	(0.187)	(0.140)	(0.213)	(0.153)	(0.214)	(0.166)	(0.263)
social_dist*active_inventor_author		-1.081***		-0.744***		-0.963***		-0.723***
		(0.289)		(0.216)		(0.276)		(0.219)
social_dist*close_paper			-0.920***	-0.728**			-0.759**	-0.572*
			(0.325)	(0.321)			(0.340)	(0.333)
N of observations (paper-patent dyads)	4,169	4,169	4,115	4,115	4,169	4,169	4,115	4,115
N of twin-patent dyads	2,068	2,068	2,041	2,041	2,068	2,068	2,041	2,041
N of citing patents	1,951	1,951	1,924	1,924	1,951	1,951	1,924	1,924
N of papers	501	501	493	493	501	501	493	493
N of twins	247	247	243	243	247	247	243	243
Log-Likelihood	-873.8	-859.0	-848.4	-842.1	-913.2	-900.1	-890.9	-884.5
R2	0.396	0.406	0.406	0.410	0.369	0.378	0.376	0.381

Notes: The unit of observation is a paper-patent dyad (n=4,169), including cited paper-patent dyads (n=2,070) and uncited paper-patent dyads (n=2,099). The sample includes 247 paper twins (7 are triplets), 501 papers, and 1,951 citing patents. We identify all cases where a patent cites only one (or a subset) of the twin papers. The binary outcome variable *cited* is equal to 1 in case the paper is cited by the patent, and 0 in case of an uncited twin paper patent dyad, i.e., the uncited twin paper matched to the same patent. The table reports coefficient estimates from conditional logit models with fixed effects at the twin patent-level. Models 1-4 include control variables: *degree3_size, no. authors, author-inventor team, publication stock, patent stock, institutional prestige, U.S. based, min geographic distance, journal impact factor, journal commercial impact factor, paper citations, patent citations, time lag, published first, text length and readability*. Models 5-8 exclude *min geographic distance* from control variables list. *Active_inventor_author* is a dummy equal to 1 if any of patent's inventors have at least 1 PubMed paper published in the year before patent filing. *Close_paper* is a dummy equal to 1, if the max similarity between inventors' prior papers (published before the patent filing year) and the twin papers (cited or not cited) is higher than sample median. Paper-patent dyads where inventors have 0 prior papers have a value of 0 for cosine similarity. Paper-patent dyads for which cosine similarity could not be calculated due to missing text data are dropped in models 3, 4, 7, 8. All control variables measured in counts are logarithmically transformed after adding one to variables with zero values. Robust standard errors clustered at the twin level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level.

Table A.11 Descriptive Statistics Comparing Corporate Inventors With and Without a Shortest-Path Connection to Cited Academic Scientists

	Corporate inventor with shortest path		Corporate inventor without shortest path		T test
	Mean	SD	Mean	SD	p value
Years since first patent/paper with same firm	4.302	4.647	3.260	4.296	0.000***
Recent public research affiliation	0.728	0.445	0.539	0.499	0.000***
Past public research affiliation	0.874	0.332	0.725	0.446	0.000***
Shared recent public research affiliation	0.251	0.434	0.150	0.357	0.000***
Shared past public research affiliation	0.373	0.484	0.200	0.400	0.000***
Number of prior patents	16.922	31.539	16.236	39.499	0.396
Number of prior patents in t-1	1.911	3.864	1.563	3.997	0.000***
Patent experience (years)	6.731	5.979	4.665	5.488	0.000***
Number of prior publications	36.733	60.572	11.189	24.820	0.000***
Number of prior publications in t-1	2.599	4.232	0.919	2.046	0.000***
Publication experience (years)	14.275	9.669	9.641	8.893	0.000***
Total network size (up to 3 degrees, log)	11.626	1.767	10.506	1.658	0.000***
Inventor network size (up to 3 degrees, log)	7.141	1.597	6.112	1.619	0.000***
Author network size (up to 3 degrees, log)	11.450	1.813	10.291	1.710	0.000***
Inventor-author network size (up to 3 degrees, log)	9.657	1.549	8.702	1.513	0.000***

Notes: The unit of observation is a corporate inventor at the time of patent filing linked to a cited twin paper. *Corporate inventor with shortest path* indicates the inventor has the shortest social distance to the authors of the cited twin paper, while *without shortest path* refers to other inventors on the same patent. *Recent public research affiliation* equals 1 if the inventor has a non-corporate (university, hospital, government, or other public research) affiliation on a paper or patent in the previous five years (t–5 to t–1); *past public research affiliation* refers to such affiliations held more than five years before filing (t–6 or earlier). *Shared recent (past) public research affiliation* equals 1 if the inventor and any author of the cited paper were affiliated with the same non-corporate organization within the recent (past) period. *Network size* variables measure the logarithm of the number of unique individuals within three degrees of separation at the time of patent filing: *total* includes all individuals, while *inventor*, *author*, and *inventor–author* network sizes restrict the network to inventors only, authors only, or those active in both domains. p-values from unequal-variance t-tests; *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A.12 Descriptive Statistics Comparing Corporate Inventors With and Without Recent Scientific Publications

	Inventors with recent publications		Inventors without recent publications		T test
	Mean	SD	Mean	SD	p value
Years since first patent/paper with same firm	4.371	4.560	3.239	4.387	0.000***
Recent public research affiliation	0.810	0.393	0.446	0.497	0.000***
Past public research affiliation	0.885	0.320	0.719	0.450	0.000***
Shared recent public research affiliation	0.287	0.453	0.109	0.312	0.000***
Shared past public research affiliation	0.361	0.480	0.227	0.419	0.000***
Number of prior patents	20.800	43.528	12.657	25.831	0.000***
Number of prior patents in t-1	1.864	3.148	1.621	4.542	0.005***
Patent experience (years)	6.484	5.968	4.983	5.609	0.000***
Number of prior publications	42.360	62.579	7.042	15.154	0.000***
Number of prior publications in t-1	3.664	4.189	0.000	0.000	0.000***
Publication experience (years)	14.122	9.687	9.765	8.936	0.000***
Total network size (up to 3 degrees, log)	11.932	1.144	10.270	1.930	0.000***
Inventor network size (up to 3 degrees, log)	7.352	1.351	5.959	1.694	0.000***
Author network size (up to 3 degrees, log)	11.768	1.171	10.044	1.985	0.000***
Inventor-author network size (up to 3 degrees, log)	9.903	1.028	8.515	1.748	0.000***

Notes: The unit of observation is a corporate inventor at the time of patent filing linked to a cited twin paper. *Corporate inventor with recent publications* indicates the inventor has at least one scientific publication in the year before patent filing (t-1), while *without recent publications* refers to inventors without recently published papers. *Recent public research affiliation* equals 1 if the inventor has a non-corporate (university, hospital, government, or other public research) affiliation on a paper or patent in the previous five years (t–5 to t–1); *past public research affiliation* refers to such affiliations held more than five years before filing (t–6 or earlier). *Shared recent (past) public research affiliation* equals 1 if the inventor and any author of the cited paper were affiliated with the same non-corporate organization within the recent (past) period. *Network size* variables measure the logarithm of the number of unique individuals within three degrees of separation at the time of patent filing: *total* includes all individuals, while *inventor*, *author*, and *inventor–author* network sizes restrict the network to inventors only, authors only, or those active in both domains. p-values from unequal-variance t-tests; *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.