Bridging Academia and Industry:
How Geographic Hubs Connect University Science and Corporate Technology

Michaël Bikard & Matt Marx*

Abstract: Innovative firms rely increasingly on academic science, yet they exploit only a small fraction of all academic discoveries. Which discoveries in academia do firms build upon? We posit that hubs play the role of bridges between academic science and corporate technology. Tracking citations from patents to about 10 million academic articles, we find that hubs facilitate the flow of academic science into corporate inventions in two ways. First, hub-based discoveries in academia are of higher quality and are more applied. Second, firms—in particular young, innovative, science-oriented ones—pay disproportionate attention to hub-based discoveries. We address concerns regarding unobserved heterogeneity by confirming the role of firms' attention to hub-based science in a set of 147 simultaneous discoveries. Importantly, hubs not only facilitate localized knowledge flows but also extend the geographic reach of academic science, attracting the attention of distant firms.

JEL codes: O00, O32

*Authorship is alphabetical. Bikard: INSEAD, michael.bikard@insead.edu; Marx: Boston University, mattmarx@bu.edu. For their valuable feedback, we thank Ajay Agrawal, Sharon Belenzon, Michael Ewens, Lee Fleming, Jeff Furman, Christopher Liu, Mark Schankerman, Eunhee Sohn, Scott Stern, Keyvan Vakili, and Ivanka Visnjic and participants at seminars at Boston University, Erasmus, the IFN, INSEAD, London School of Economics, the Stockholm School of Economics, UC Berkeley, and Wharton and at the Duke Strategy Conference, Georgia Tech Roundtable for Engineering Entrepreneurship Research, Israel Strategy Conference, West Coast Research Symposium, and NBER Productivity Lunch and Summer Institute. We thank Pierre Azoulay for sharing his crosswalk between PubMed and the Web of Science. This work was supported by a Kauffman Junior Faculty Fellowship, National Science Foundation Science of Science Policy Grant #1735669, and by the Institute of Innovation and Entrepreneurship at London Business School.

Introduction

Fundamental scientific discoveries are essential to increasing productivity and innovation (Nelson 1982; Rosenberg and Nelson 1994; Mokyr 2002; Fleming and Sorenson 2004), resulting in "new products, new industries, new investment opportunities, and millions of jobs" (Bush 1945, 242). Yet firms are funding less and less internal R&D, relying instead on discoveries from academia (Arora, Belenzon, and Patacconi 2018). Considering this division of innovative labor (Rosenberg and Nelson 1994), firms exploit a surprisingly small fraction of the work published by academic scientists. Building on academic science is a challenging task (Bikard 2018), and less than 5% of all academic publications are cited by patents (Ahmadpoor and Jones 2017). This raises a key question regarding the interface between academic science and industrial technology: Which scientific discoveries from academia do firms choose to build upon?

Explaining why some academic discoveries attract the attention of firms while others do not is of fundamental strategic importance as ability to identify and exploit useful scientific knowledge may confer considerable competitive advantage. Consider the case of the purification of erythropoietin (i.e., EPO) by Eugene Goldwasser, a biochemist at the University of Chicago. The pharmaceutical industry initially ignored this breakthrough, even though he reached out to various firms. Eventually, Amgen paid attention to Goldwasser's findings, allowing it to secure valuable intellectual property, and after a few years of development, used Goldwasser's insights to develop Epogen, one of the most successful drugs ever (Goldwasser 2011).

In this paper, we highlight the role of *hubs* of industrial R&D as an interface between academic science and industrial technologies. By hubs we mean a geographic concentration of patenting by firms in a specialized technical field.[1] We posit that hubs facilitate the flow of knowledge from academia to industry. Our initial finding (summarized in Figure 1) – that academic publications from hubs are much more often cited in firms' patents – is in line with prior research on the benefits of collocation for knowledge flow between universities and firms (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Gittelman 2007; Belenzon and Schankerman 2013). However, Figure 1 also shows that the vast majority of the knowledge flows from hub-based published academic research are *not* localized. Rather, distant firms also pay more attention to academic discoveries in hubs, extending the geographic reach of academic science.

[1] For example, an article on semiconductor fabrication published in Dallas, TX would be in a hub as there are many nearby semiconductor firms, whereas a semiconductor article published in Boston, MA would not.

Figure 1 about here

Our empirical approach involves two complementary analyses. First, we examined citations to over ten million journal articles in the Web of Science in patents granted to firms. We found that although only a small fraction of articles came from a hub, they were cited far more frequently in firms' patents; that hub-based academic science was of higher quality, accruing more citations in academic papers and appearing in journals that had higher impact factors; and was more applied. We measured "appliedness," introducing the construct *journal commercial impact factor (JCIF)* and finding that hub-based papers were published in more applied journals and fields. Taking this into account, hub-based discoveries were twice as likely to be cited in firms' patents.

A number of patterns emerged from this to suggest that firms pay more attention to papers from hubs. The hub effect is 1) stronger for newer discoveries, when uncertainty is highest; 2) weaker for academic discoveries involving collaboration with industry; 3) stronger when hubs are specialized in a few technologies; 4) stronger among firms that are younger, more innovative, and more reliant on science. It should be pointed out, however, that the cross-sectional nature of this large-sample analysis does not allow us to rule out unobserved characteristics that might otherwise explain the hub effect.

We take advantage of simultaneous discoveries (Merton 1961) to test more cleanly whether firms pay more attention to hub-based science. When multiple scholars publish the same findings, they create "paper twins" (Bikard 2012). We use 147 simultaneous discoveries published in 316 academic articles. Indeed, we find that firms are more likely to cite the hub-based member of a paper twin, even when controlling for paper-patent proximity. Importantly, a hub-based paper twin is not presented in a more "applied" manner. Neither do explicit social networks such as co-authorship explain this effect, although it may be partially driven by informal interactions between academic and industry scientists as the hub effect is stronger for hubs that host more Gordon Research Conferences (which are attended by both industry and academic scientists).

Our study makes two main contributions. First, we show that hubs serve to build cross-institutional bridges, facilitating the flow of academic science to industry. Second, our results show that this bridging role of hubs extends beyond the frequently-studied localization of knowledge flows (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Gittelman 2007; Belenzon and Schankerman 2013; Singh and Marx 2013), attracting the attention of distant firms to local academic discoveries. To promote future work on this topic, the list of hubs for each patent technology subclass is available to download, as is our journal commercial impact factor measure.

1. Hubs as an interface between academic science and industrial technologies

1.1. Firms' exploitation of academic science

Science boosts firms' innovation performance not only by increasing the productivity of their existing R&D work but also by inspiring entirely new projects (Nelson 1982; Mokyr 2002; Cohen, Nelson, and Walsh 2002; Fleming and Sorenson 2004). Notwithstanding, scientific inquiry is largely the preserve of academic institutions rather than firms – there is a distinct 'division of innovative labor' between those who explore and exploit scientific knowledge (Bush 1945; Rosenberg and Nelson 1994; Arora and Gambardella 1994). In the past quarter century, firms have steadily withdrawn from internal scientific inquiry, as documented by Arora, Belenzon, and Patacconi (2018). Their examination of 4,608 companies publicly-listed in the US between 1980 and 2006 found that the average number of scientific articles authored annually by those firms fell by 20% per decade. Importantly, there has been no parallel decline in innovative output – simply an increasing reliance on knowledge from academia (see also Bena and Li (2014).

This reliance on academia relieves managers of the need to invest in uncertain R&D, yet it comes with two challenges. The first relates to the supply of academic publications that can be usefully exploited by firms. The primary goal of scientists in academia is to advance understanding rather than advance technology. Indeed the tension between the two objectives is well-documented (Gittelman and Kogut 2003; Stern 2004; Murray 2010; Bikard 2018). So it should come as no surprise that only a small fraction of academic science is directly useful for industrial innovation. Rosenberg and Nelson (1994) note that two thirds of research funding in the US is directed at basic research—albeit difficult to distinguish between basic and applied work (see also Stokes 1997). Ahmadpoor and Jones (2017) examine the relevance of academic science for technology development by measuring the minimum citation distance between scientific advances and follow-on patented inventions. In a sample that included 4.8 million US patents and 32 million research articles, only 1.41 million articles were cited in the patent literature. However, when indirect connections were considered—scientific articles were cited by other articles, eventually linking to a patent citation—79.7% of the science and engineering literature could ultimately be linked to a patent. Hence the link between science and technology is primarily indirect.

Moreover, publications by university researchers are often of poor quality. The quality of academic literature is marred by concerns about integrity and replicability. The rate of replicability of published academic studies is estimated to be between 11% and 49% (Begley and Ellis 2012; Hartshorne and Schachner 2012; Freedman, Cockburn, and Simcoe 2015). For firms, this adds considerable uncertainty and increases the cost of science-based technology development (Osherovich 2011; Rosenblatt 2016).

A second challenge for firms is the sheer size of the academic literature and its speed of evolution. Since the first scientific article was published in the *Journal des Sçavans* in Paris on January 5, 1665, over 50 million manuscripts have been published. The speed at which new contributions are added is increasing (Jinha 2010), adding to the burden on researchers who must lengthen their training, specialize, and collaborate in order to stay up to date (Jones 2009; Agrawal, Goldfarb, and Teodoridis 2016). Inventors in firms face an even larger burden given their limited involvement with academia (Agrawal and Henderson 2002; Murray 2002). Clearly, firms that seek to exploit academic discoveries risk wasting their efforts in the wrong corner of the literature, while those that ignore academic work may miss out on opportunities to improve their R&D efficiency.

The decision to pay attention only to certain academic publications is therefore a strategic one. Simon (1947: 124) noted: "In the contemporary world all of us are surrounded by, even drowned in, a sea of information, only an infinitesimal part of which can be attended to. Although we may wish to have certain kinds of information that are not available (e.g., reliable forecasts), the critical scarce factor in decision-making is not information but attention. What we attend to, by plan or by chance, is a major determinant of our decisions." This allocation of attention in turn shapes the direction and efficiency of innovative efforts and has fundamental implications for firm performance (Ocasio 1997, 2010). Indeed, there is evidence that researchers' scarce attention to new knowledge shapes the evolution of science (Simcoe and Waguespack 2011; Iaria, Schwarz, and Waldinger 2018; Teodoridis, Bikard, and Vakili 2018; Reschke, Azoulay, and Stuart 2018; Chai and Menon 2019) while R&D workers' bounded attention to new inventions shapes the development of new technologies (Podolny, Stuart, and Hannan 1996; Polidoro 2013; Drivas, Lei, and Wright 2017). However, much less is known about the interface between the two—i.e., how firms allocate their limited attention to the academic literature. This gap in our understanding is surprising considering that firms primarily learn about the work of academics by reading their published research (Agrawal and Henderson 2002; Cohen, Nelson, and Walsh 2002).

1.2. Hubs as cross-institutional bridges

We argue that hubs serve as cross-institutional bridges, fostering the flow of academic science into industrial technologies, and see two main drivers of this phenomenon.

The nature of academic science in hubs. We posit that hub-based academic discoveries are more useful from a technological standpoint than that originate elsewhere. The uneven distribution of technologically valuable academic discoveries may be explained by the fact that academics who value commercialization are attracted to hubs because collocation with relevant firms may present career opportunities. Indeed innovative firms often cluster around academic stars (Zucker, Darby, and Brewer

1998). Industry proximity may boost academics' productivity and the quality of their work. Interactions between the two may encourage learning, cross-fertilization of ideas, collaboration, and open up access to a firms' equipment, capabilities, and financial resources (D'Este and Perkmann 2011; Chai and Shih 2016; Bikard, Vakili, and Teodoridis 2019). Firms might even directly influence the direction of academic work to topics that they have an interest in (Furman and MacGarvie 2007; Evans 2010; Sohn 2014). For these reasons, hub-based academic work will, on average, be of higher quality and have higher value for technology development than academic knowledge produced elsewhere.

Firms' attention to academic science emerging from hubs. As firms' attention is limited, it will be allocated to some scientific papers but not others. Hubs are likely to attract firms' attention for at least two reasons. First, if firms know that hub-sourced academic knowledge is more useful from a technological standpoint, they are naturally drawn to those discoveries; the hub serves as a cue for firm attention, in much the same way as the social status of the author signals that an article deserves attention (Merton 1968; Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2013). Second, since firms are generally more exposed to academic science from hubs, this relative difference in exposure may lead to systematic divergence in the amount of attention received. Firms based in hubs are naturally more exposed to local academic work (Jaffe, Trajtenberg, and Henderson 1993; Belenzon and Schankerman 2013). Hubs tend to foster frequent interactions, which play an important role at the intersection between academia and industry (Cohen, Nelson, and Walsh 2002; Gittelman 2007). However, we posit that firms located *outside* a hub are systematically more exposed to academic knowledge that comes from hubs, as inventors often travel to hubs or through hubs, where they hear about the work of local academics.

1.3. Patent-to-paper citations and science-based invention

To understand how employees exploit the scientific commons, a survey by Cohen, Nelson, and Walsh (2002) asked industry inventors which pieces of academic knowledge they relied on in their innovation activities. Since this may exceed the inventor's power of recall, especially for work performed in the distant past, we take a different approach, relying on the fact that applicants for patents are required to list the academic literature used for their invention. While it is possible that inventors consulted but did not cite a paper, the list is likely to constitute a useful "paper trail" of the science inventors utilized at the time.

We follow a small but growing literature that has used patent-to-paper references as a proxy for the flow of scientific knowledge into new technologies (e.g., Cassiman, Veugelers, and Zuniga 2008; Belenzon and Schankerman 2013; D. Li, Azoulay, and Sampat 2017; Ahmadpoor and Jones 2017). Using this measure is supported by Roach and Cohen's survey finding that "citations to nonpatent references,

such as scientific journal articles, correspond more closely to managers' reports of the use of public research than do the more commonly employed citations to patent references" (Roach and Cohen 2013, 505). However, more detailed fieldwork reveals this to be a noisy measure of knowledge flow because it captures some citations that are not directly related to the invention, while failing to capture others that are (Callaert, Pellens, and Looy 2014). This raises a question about the process of citations of academic literature in the patenting process: How do inventors determine which publication to cite in their patent when more than one seems appropriate?

To understand the "data generation process" underlying patent-to-paper citations, we held a series of 21 interviews with inventors (Appendix A1 describes interviewee selection), from which it was clear that patent-to-paper citations primarily reflect industrial inventors' awareness of the academic literature. In line with prior studies on this topic (Lemley and Sampat 2011), interviewees confirmed that their citations of the academic literature came from them (not from an attorney or patent examiner), but also recognized that the list of citations was not exhaustive – they simply wrote down what they recalled at the time of the patent application. Said one inventor, "*Scientists are not as careful for patents as they are for their own publications.*" Another said : "*You know, it is sort of an abbreviated list of citations that establish a certain sense of what's known in the art but it's not necessarily exhaustive or for that matter even fair in terms of giving credit*." Thus their patent-to-paper references reflected not only the relevance of a scientific publication to an invention but the attention they paid to it, as illustrated by one inventor interviewed: "*The more familiar paper comes to mind first, plus the citation makes the point - done! I think it is convenient to say that we know about all the other papers, but there might be examples where in fact we don't know. I can't say that it is always true that we were aware of everything.*" Thus inventors cite publications to which they give most attention.

From this we postulate that hubs build cross-institutional bridges that channel the flow of knowledge from academia to industrial technology. Like prior research, our (admittedly small-scale) fieldwork suggests that patent-to-paper references measure this knowledge flow by making it possible to observe which academic publications firms read – hence our hypothesis that academic publications from hubs receive a disproportionate number of citations from corporate inventors.

2. Large-sample analysis

Our first step was to examine more than 50 million journal articles from the Web of Science (WoS) for the years 1955-2017. Fields include year, title, journal (including volume, issue, and page number), first author's name and institution (along with institution's address), and digital object identifier.

Since our interest is whether firms pay attention to academic science, our first step was to reduce the

set to academic articles only. The corresponding author's institution was available in about 45 million articles. Institution names are entered in World of Science (WoS) as free text (no unique identifiers) so we manually reviewed tens of thousands of institution names comprising more than 25% of citation-weighted papers. We classified each institution as academic, government, or industry using rule-based systems and manually reviewed every institution with at least 250 articles. Of the WoS articles for which we had an institutional affiliation, 91.2% were from academic institutions.

We generated several measures of quality. First, WoS includes a broad 'subject' field for each paper, 251 fields in total. For each institution, we counted the number of papers in the same field as the focal paper as a measure of that institution's prestige in that subfield. The same was done for authors. At the paper level, we took advantage of the fact that WoS reports nearly one billion article-to-article citations. We used these citations to calculate both journal impact factor (JIF) and individual article quality (the latter computed as the number of forward citations up to five years after publication date).

2.1. Dependent variable: patent-to-paper citations

Determining whether firms pay attention to academic science is challenging. A firm is a collection of individuals, so reliably tracking whether each employee paid attention to every relevant piece of scientific knowledge would involve a daunting quantity of fieldwork. However, for innovative employees—namely, patenting inventors—the academic works they cite in their own patents serve as a proxy for their attention. WoS does not record patent citations to scientific articles, so we used the patent-to-paper citations dataset constructed by Fleming et al. (2017), which contains a full explanation of their algorithm. For convenience, we summarize the approach below.

Tracking citations to papers is more difficult than to patents because patent-to-paper references are listed as unstructured strings. Instead, we scanned for article-related fields from WoS in the unstructured references listed on the front page of the patent documents under "Other References." Titles are frequently abbreviated or misspelled, which we attempt to account for using fuzzy matching. Indeed, titles are often missing entirely or abbreviated so aggressively that even fuzzy matching methods fail. Thus, as a second pass, we searched for volume-issue-page in sequence without any other numbers in between, on the assumption that that {year, author, volume, issue, first page} would be unique across the papers. This exercise yielded approximately 9MM patent-to-paper references across the entire Web of Science since 1955. Known limitations of the algorithm include misspellings of the author's surname in the unstructured patent reference and an incorrect (or missing) year in the reference.

2.2. Defining "hubs" of industry R&D

Our analysis turns on whether a given paper is in a "hub" – i.e., a geographic concentration of industrial R&D relevant to the paper. To determine hubs, we start by collecting the technological subclassifications from all patents, whether industry or academic, that cite at least one scientific article in order to have the most complete possible representation of USPTO patent subclasses that are applicable to the discovery. For each subclass we collect all firm-assigned patents belonging to that subclass (whether or not they reference any scientific article).

For each subclass we count the number of industry patents for each half-decade and also the number in each city during that same time window. We divide the latter by the former to obtain the percentage of patenting of that subclass in that city during that five-year period. A city is classified as a "hotspot" of patenting for a given subclass if it satisfies two criteria. First, >5% of patents in that subclass must be in that city. Second, because this threshold can easily be exceeded in subclasses with few patents (e.g., in a subclass with only 20 patents, *every* location has 5% of patenting), a city must have at least five patents in that subclass to qualify. Table A1 shows the hotspots for 20 randomly-sampled subclasses. Figure A2 provides a world map for hotspots of all subclasses.

We then form a "hub" with a hotspot city at the center of a circle with a 50-mile radius—approximating reasonable commuting distance. Our *Paper in a hub* variable reports whether the corresponding author is in a hub for any of the paper's associated subclasses (in other words, whether that author is within 50 miles of any of the paper's subclass hotspot cities). To do so, we need to match articles to patent subclasses. For papers cited by patents, the technical classification(s) of the citing patent(s) are used. For papers not cited by patents, we collect relevant patent subclasses from articles in the same field. Because the Web of Science uses very general field descriptors (n=251), every paper in broad fields (e.g., Molecular Biology or Mechanical Engineering) would have the same hubs even though the papers are on widely varying topics. Instead, we leverage two other sources to determine more precisely the subclass a given paper belongs to. The first source of detailed topics is PubMed, which reports any the of 27,255 MeSH keywords associated with approximately 13 million WoS articles.[2] The second is the Microsoft Academic Graph (MAG), which covers 166 million articles. MAG automatically generates topics for each paper by processing the text of the title and abstract, with 185,600 fine-grained topics. We construct a crosswalk between approximately 23 million WoS articles and MAG based on 1)

[2] We thank a reviewer for noting that MeSH keyword classifications have evolved over time, so our results may be sensitive to such changes. We therefore re-estimated our main effects only using papers published since 2012 and found similar results.

exact DOI matches; 2) exact title matches where the year and either the volume, issue, or first page number also match; 3) "near" matches of titles with more than 15 characters where the year as well as two of year, issue, and first page match and the Levenshtein distance between the titles is less than 2% of the title's length.

For each MeSH or MAG keyword, we collect the technology subclasses for all patents that cite any article with that keyword and rank subclasses within keyword by the number of articles with which that subclass is associated. We discard all but the most popular patent subclasses for each keyword, retaining ties. However, if even the most popular patent subclass(es) for a keyword are not associated with at least 25% of the papers in that keyword, these are also discarded as they may not truly be representative of the keyword. For instance, if a patent subclass is the most common for a keyword with 1,000 papers but is used only by a patent citing six of those 1,000 papers, it is deemed to be unrepresentative of the keyword. Many keywords are not well-represented by any patent subclasses; papers whose keywords are not well-represented by at least one patent subclass cannot be evaluated as to whether they are in hubs.

This algorithm enables us to attach one or more patent subclasses to any article that 1) itself was cited by a patent, or if not, 2) had a keyword where other papers with that same keyword were cited by patents and where one or more technology subclasses were clearly dominant. We find this to be the case for 10.2MM WoS papers, so we label each of these papers as in a hub or not. For the other papers, either we could not find MeSH/MAG keywords, or fewer than 25% of the papers with those keywords were cited by patents in a particular field. On average, the papers for which we can assign patent classes appear to be of higher quality and more applied than those for which we cannot. Of those we can assign patent classes, 3.8% of articles are in hubs. The latitude/longitude used to define each hub can be downloaded at https://archive.org/details/hubstopost_201903.

2.3. Applied vs. basic orientation of science

One factor that may affect the propensity for an article to be referenced by a patent is the degree to which it represents applied science. We build two measures of appliedness, as follows.

Appliedness of paper: To calculate the "appliedness" of each article we turn once more to MeSH keywords and MAG topics. For each combination of MeSH and MAG keywords, we sum the number of patent-to-paper references for papers with the same combination of keywords. This count of five -year citations *from patents* at the level of the keyword combinations yields a measure of how "applied" this paper's specific field is. As an example, instead of comparing a focal paper with all others in Molecular Biology field from WoS, we compare it with papers also using the MeSH keywords Mitochondrial DNA, Birds, and Restriction Fragment Length Polymorphism. Of the 10.2MM papers for which we are able to

calculate hubs, we find either PubMed or MAG keywords for 9.8MM of those and thus are able to create appliedness measures. This measure is a fine-grained one since it is computed at the keyword-combination level, but as an illustration we also compiled the score at the level of Web of Science fields. Table A2 shows the twenty most applied fields, the top three being Computer Science Hardware & Architecture, Software Engineering, and Medicinal Chemistry.

Journal Commercial Impact Factor (JCIF): This measure is analogous to the commonly used Journal Impact Factor (JIF), which is a popular measure of its quality, calculated for year *t* as the number of times articles from years *t-1* and *t-2* were cited during year *t*, divided by the number of articles published during years *t-1* and *t-2*. Just as JIF is a journal-level measure of quality, it is possible to build a journal-level measure of appliedness by replacing paper citations to papers by patent citations to papers. We also publish these data for use by other researchers: https://github.com/mattmarx/jcif.

2.4. Empirical setup and results

Of the WoS papers for which we are able to determine hubs, approximately 10.2MM are published by academic institutions and are the object of our examination. Descriptive statistics are in Table I. Approximately a third of those papers are from the United States. Papers have on average 15.1 citations from papers within five years of publications but barely 0.06 citations on average from industry patents. Importantly, Table 1 also shows that only 3.8% of academic articles stem from hubs.

Table I about here

Our investigation of industry patents citing academic articles begins Table II, which shows difference of means tests according to whether the focal article is in a hub. Although in Table I we saw that few academic articles emerge from hubs, Table II shows that hub-based articles receive on average 4.6 times more citations from firm-assigned patents than articles outside hubs in the five years following their publication. This very large difference is also visible at the firm level. For example, the 3.8% of the papers that are in hubs represent 25.4% of the academic articles cited by the three firms with the most patent-to-paper citations.[3] Table II also shows that hub and non-hub papers are quite different. Hub papers appear to be of higher quality (81% more paper citations within five years, 48% higher JIF) but also much more applied (28% higher paper appliedness, 115% higher JCIF).

[3] Those firms are Genentech, Microsoft and IBM. The proclivity of firms whose patents cite articles to focus on hub-based articles extends broadly. Among the top 500 firms, 20.4% of their citations are to hub-based articles. This figure is 18.2% among the top 10,000 firms.

Table II about here

The core of our large sample analysis is presented in Table III. Again, we find that the difference in firm patent citations to hub- and non-hub academic papers is considerable. Controlling for the year of publication and US origin in Model 1, the incidence rate ratio of citations from industry patents within five years of publication is 3.4 times higher for academic papers in hubs. One might suppose that this effect is driven by the geographic proximity of hub-based academic scientists to hub-based firms. Indeed, a vast literature has emphasized the crucial role of collocation for knowledge flow between academia and industry (Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998; Gittelman 2007; Belenzon and Schankerman 2013). Yet the hub effect is *not* explained solely by the localization of knowledge flows, as suggested by Figure II. In Model 2 we restrict the dependent variable to count only citations from patents located more than 50 miles away (i.e., outside the hub). If the hub effect were merely epiphenomenal with knowledge localization, we would expect the estimated coefficient on *Paper in a hub* in Model 2 to be substantially weaker, yet it retains both the magnitude and statistical significance from Model 1. This indicates that papers in hubs of relevant commercial R&D attract attention not just from local firms but from distant ones as well.

We theorized that the apparent effect of hubs as bridges between academic science and industrial technologies will be driven in part by the nature of hub-based academic science, Model 3 introduces measures of quality and appliedness, both at the level of the paper and at the level of the journal in which it is published. Our paper-level measures of quality and appliedness are strong positive predictors of citation in firms' patents. At the journal level, however, we find that JIF is a negative predictor, whereas JCIF is a positive predictor of citations in firms' patents. This is perhaps not surprising given that firms should value commercial relevance over academic renown. Model 3 shows that those measures of the nature of academic science account for a considerable share of the hub effect, reducing its incidence rate ratio from 3.4 to 2.1. In Model 4, accounting for the prestige of the corresponding author's institution and the collective prestige of the authors—in the subject area of the focal paper as captured by the number of publications in that subject area—does not shift the coefficient estimate materially. After accounting for the quality and appliedness of hub-based science, we find that hub location is associated with a 112% boost in citation in firms' patents.

We also theorized that firms pay more attention to discoveries located in hubs of industrial R&D in relevant fields, Models 5-7 start our exploration of search heuristics. First, we explore the role of academia-industry collaborations, defined as a paper where the corresponding author is at an academic institution but (at least) one non-corresponding author is from industry. About 9% of our 10.2 million papers include such collaborations. Industrial origin is known to attract inventors' attention (Bikard 2018),

hence we should expect that those collaborative papers will attract more attention from industry inventors. In line with this idea, Model 4 shows that academic papers that include industrial coauthors are more highly cited in firms' patents. We also find that the interaction between university-industry collaboration and hub location is negative and statistically significant, suggesting that academic collaborations with industry and presence in hubs serve as substitutes.

Table III about here

Second, we distinguish between different types of hubs. If attention plays a role, we would expect specialist hubs to get more attention than generalist hubs. The benefits of specialization for visibility have long been highlighted in sociology (e.g., Zuckerman and Kim 2003; Leahey 2007). In Model 4, we therefore split the hub variable between "specialist" and "generalist hubs." To identify specialized vs. diversified hubs, we calculated the Hirsch-Herfindahl Index based on the number of USPTO technology classes for which that location served as a hub. If that location was a hub for only a single technology class, its "hub HHI" would be 1. We then split the "hub" variable into two variables indicating that the paper is in a hub with HHI a) greater than or equal to 90% of all other hubs, b) lower than 90% of all other hubs. In Model 6, we find that the estimated coefficient for specialized hubs is about double that of diversified hubs (joint significance $p<.001$). The results for diversified vs. specialized hubs are, moreover, robust to a 75/25 cutoff, or comparing hubs with HHI of 1 (i.e., only a single technology class represented in the hub) vs. all others. Estimates are less precise using a median split.

Third, we consider academic inventors. If attention plays a role, we would expect the estimated coefficient for hubs to be smaller for citations from patents assigned to universities. Academic inventors will be more familiar with the academic literature than corporate inventors and should allocate their attention in a way that relies less on hubs. Indeed, we can see by comparing Models 4 and 7 that the hub coefficient is much smaller for academic inventors than for corporate ones (0.462 vs. 0.752).

In addition, we explore the impact of uncertainty. Since we theorize that firms use hubs to resolve uncertainty about the usefulness of a discovery, we expect reliance on hubs to decline as uncertainty decreases. Clearly, uncertainty is highest at the knowledge frontier. Over time, follow-on studies and replications clarify the value of the knowledge on which they build, just as articles become better understood over time, accruing citations from other articles which serve as a signal of quality. It follows that hubs are likely to be more useful heuristics at the scientific frontier than regarding older findings; put another way, firms' reliance on hubs decrease as the science ages. Figure 2 explores this relationship and shows that reliance on hubs is highest for publications that are less than five years old. After that, the share of firms' citation to hub papers in their patents declines sharply.

Figure II about here

Table IV further examines the type of firms that make use of hubs. To do so, we shift the unit of analysis and explore the correlation between firm characteristics and the extent to which they cite hub-based science. For the 36,769 firms whose assigned patents cite any of our approximately 10 million WoS papers, we examine the ratio of each firm's citations that are to papers in hubs. (Descriptive statistics are in Table A3.) Since we theorize that hubs act as bridges between academic science and corporate technologies, we would expect firms that draw on hub science to be particularly innovative and science-oriented. In practice, we find that firms that draw on hubs tend to be younger (Model 1); that they are more innovative as measured by their patent stock (Model 3) and forward citations (Model 5); and that they are particularly reliant on science as measured by their average number of patent to paper citations (Model 7). These relationships hold jointly in Model 9; moreover, when adding fixed-effects for the firm's most common patent subclass, the results are preserved in Models 2, 4, 6, 8, and 10.

Table IV about here

In sum, our large-sample analysis highlights the role of hubs as bridges between academic science and industrial technology. Few academic publications emerge from hubs, but those that do are disproportionately cited in firms' patents. In part this is because academic science from hubs is of higher quality and more applied. However, we also find evidence suggesting that firms pay particular attention to academic science that emerges within hubs. Our results indicate that the firms that rely on hubs the most tend to be innovative, and that firms use hubs to stay at the forefront of the fast-evolving scientific discoveries.

Our large-sample analysis is not without limitations. In particular, individual scientific discoveries may be more or less attractive to firms in ways that are difficult to capture via the quality or appliedness measures we have constructed. Some of our measures may even blur the line between inherent quality and attention. For example, papers might be highly cited in the academic literature because they are of high quality or received a lot of media coverage, which would also attract attention from inventors in firms. In what follows, we seek to address this concern by adopting a complementary empirical strategy that enables us to more fully control for the nature of the scientific discovery.

3. Twin-paper analysis

3.1. Dataset

For our second analysis, we use simultaneous discoveries in the natural sciences, that is, instances when the same or very similar knowledge emerges in multiple locations. These "paper twins" present a

unique opportunity to unbundle producers from their products, and allow us to measure firms' attention directly by examining patent references to the academic publications which make up each set of twins while accounting for the characteristics of the individual scientists and of their institution.

Our study is based on the first systematically collected dataset of simultaneous discoveries. The nature of a simultaneous discovery is illustrated by the following example.

The August 1998 issue of *Cell* contains two papers (Figure III) reporting the discovery of an important molecule involved in cell death or "apoptosis." The two teams found that after activation of the death receptors on the cell membrane, the death signal is carried to the mitochondria by a cytosolic protein called BID. Confirming that these two papers truly report the same scientific discovery, an August 21 2000 article in *The Scientist* notes that "[t]hese two *Cell* papers outline two independent identifications of a critical missing link in [the apoptosis] signaling pathway" (Halim 2000). Frequently, in the case of simultaneous discoveries, authors send their manuscripts to the same journal, sometimes leading to back-to-back publications[4] (in this case: pages 481-490 and 491-501; in our dataset, 46% of the simultaneous discoveries are published back-to-back in the same issue of the same journal).

Figure III about here

To detect simultaneous discoveries, an algorithm was built that identified frequently co-cited pairs of papers and then we scrolled through the scientific literature to spot instances in which two papers are consistently cited *in the same parenthesis*, or adjacently. The method is explained in detail in Bikard (2012). We present a summary below for convenience.

Sociologists of science have found that citations provide a window into the scientific community's allocation of credit (Cozzens 1989). The algorithm therefore considers pairs of scientific publications that are consistently cited together—i.e., in the same parenthesis, or adjacently. The algorithm proceeds as follows. A dataset consisting of 42,106 scientific articles published between 2000 and 2010 was built from the 15 non-review scientific journals having the highest impact factor in 2009. Three-quarters of a million unique references from these articles were grouped into pairs that (a) were co-cited at least once, (b) were written no more than a calendar year apart, (c) have no overlapping authors, (d) at least 5 citations for each reference are observed. A Jaccard co-citation coefficient—i.e., ratio of the intersection over the union of forward citations—was then calculated for all pairs meeting these criteria; pairs with a Jaccard

[4] Editors sometimes publish manuscripts back-to-back, recognizing a tie in the race for priority and allowing both teams to receive equal credit for their work. For example, Darwin and Wallace published their theory of evolution by natural selection back-to-back in the *Journal of the Proceedings of the Linnean Society of London* on 20 August 1858.

coefficient >50% were retained. Finally, those pairs for which 100% of the co-citations took place in the same parenthesis or were retained adjacently. These 720 pairings of 1,246 papers disclosed 578 simultaneous discoveries since there are instances of discoveries involving three or more teams.

We discarded 99 papers published by firms or whose only twin was a firm-published article, as our aim is to understand which academic publication firms focus on, leaving 1,147 papers. We formed simultaneous-discovery/patent dyads to analyze whether a given patent cites one paper but not its twin(s). For 831 of the papers, in all of the simultaneous-discovery/patent dyads to which they belong, the patent either cited none or all of the twins. Since there is no variation in the dependent variable for these twin papers, they are dropped from our models, which utilized fixed effects for each simultaneous-discovery/twin dyad. This leaves 316 papers. (However, our results are robust to using linear probability models, which include instances in which patents cite all the papers of a set of twins).

For each publication, we noted its geographic origin, journal, and whether the discovery was itself patented. (Whether the focal paper reporting a simultaneous discovery was patented by its authors is essential to form a "patent paper pair" (Murray 2002).) To account for author heterogeneity, we collected the corresponding author's stock of patents and papers at the time of publication. Similarly, we captured the institution's stock of patents (past five years). As a measure of the institution's prestige in the particular field of that paper, we counted the number of papers ever published by that institution in the top 15 scientific journals relevant to that paper, as calculated from Medical Subject Heading (MeSH) keywords. For each article, we selected the ISI classification most frequently associated with its MESH keywords.[5]

Table A4 in our Appendix provides a breakdown of the most frequent cities and institutions among the academic publications in our analysis of paper twins, where one but not all papers of a set of paper twins were referenced by some patent, as well as the most frequently referencing patent assignees. As seen in Panel A, many twin papers stem from high-status institutions, including nearly 5% from Harvard University. This underscores the importance of accounting for the prestige of the institution (since industry inventors may be more likely to be exposed to such discoveries). A similar concern also applies in Panel B, which shows that nearly 20% of all twin papers are published in Boston, New York, and San Diego. Panel C shows, unsurprisingly, concentration among patent assignees, although our methodology

[5] The ISI keywords for our journals are Biochemistry & Molecular Biology, Biotechnology & Applied Microbiology, Cell & Tissue Engineering, Cell Biology, Physical Chemistry, Genetics & Heredity Immunology, Materials Science, General & Internal Medicine, Research & Experimental Medicine, Nanotechnology, Oncology, Applied Physics, Condensed Matter Physics, and Interdisciplinary.

of assessing whether a particular patent cites one twin paper vs. another helps to ameliorate this concern.

As with our large-sample analysis above, we measure firms' attention to science by observing their patent citations to academic publications—this time focusing on paper twins. For this analysis, we checked the patent citations to twin papers and categorized the patents as academic or firm-assigned patents by hand. Since paper twins disclose the same discovery, each paper is admittedly as cite-able from the inventor's standpoint. Even though those papers are published and (presumably) equally relevant, boundedly rational inventors may not have paid equal attention to every paper of a set of paper twins. It is important to note that the U.S. Patent and Trademark Office imposes no duty on the inventor to reference *every* paper disclosing the same simultaneous discovery. According to USPTO Rule 56 (37 CFR 1.56): "information is material to patentability when it is not cumulative to information already of record or being made of record in the application." In other words, if multiple papers disclose the same knowledge, referring to just one is sufficient. In line with this rule, our interviewees highlighted that they did not attempt to be comprehensive in their patent references to the academic literature.

3.2. Empirical Setup

Our analysis leverages the simultaneous-discovery nature of our data because a patent that references one paper is presumably at a similar risk of referencing its twin. An observation is a dyad of a published paper reporting a simultaneous discovery and a patent at risk of referencing the paper. To obtain a dataset that includes not only realized patent-to-paper citations but also unrealized citations to the twins of a focal paper, we pair each industry patent[6] that references a paper with the other "twin" papers that disclose the same simultaneous discovery. That is, given a set of paper twins where one is referenced by a later patent, we create an observation for that same patent together with the twin paper that was not referenced. This process yields 1,671 paper-patent dyads for 316 sets of twin papers and 697 patents.

For each paper-patent dyad representing a (potential) scientific reference, we account for both temporal and spatial separation between the paper and the patent. Given that our explanatory variable reports whether a paper is located in a hub, distance between the paper and the potentially-referencing patent is an important control. As noted above in our large-sample analysis, one might conjecture that papers in hubs are simply within shorter distances of potentially-citing patents and therefore more likely to be cited. Thus all of our results using the twin-paper analysis accounts for distance in a linear fashion

[6] Twin papers are cited by three general types of patents: those assigned to firms, those assigned to academic institutions, and unassigned patents which remain the property of the inventor(s). We reviewed every patent that cited any of the twin papers to determine whether it was firm-assigned, university-assigned, or unassigned. Only firm-assigned patents are included in our principal analysis, although we use university-assigned patents as a placebo test. Unassigned patents do not enter into our estimations.

with the logged count of miles between the paper and potentially-referencing patent. Also, twin papers are usually (but not always) published in the same calendar year, so we control for the lag between the publication of the twin and the potentially-referencing patent. Summary statistics are in Table A5.

We estimate a conditional logit model with fixed effects for the simultaneous discovery and a focal patent that references one but not all papers of the set of paper twins reporting that discovery. Thus, for a given patent that is arguably at equal risk of referencing any of the paper twins, our analysis reveals the factors associated with a particular twin being referenced. The regression equation is given as

$$REFERENCED_{ijk} = f(\varepsilon_{ijk};\ \alpha_0 + \alpha_1 IN\_HUB_i + \alpha_{2\circ}\bar{X}_{ik} + \gamma_{jk})$$

where *j* represents the simultaneous discovery, *i* represents a paper reporting the simultaneous discovery, and *k* represents the potentially-referencing patent. IN_HUB_i is the main explanatory variable and is defined at the paper-patent dyad level. γ_{jk} is a simultaneous-discovery/patent fixed effect, which allows us to be unconcerned with within-firm drivers of attention and to focus instead on external ones.[7] Finally, X_{ik} is a vector of covariates including the geographic distance between the focal paper and the potentially-referencing patent. Standard errors are clustered at the level of the set of paper twins.

3.3. Paper Twin Analysis: Main Results

We begin our analysis in Table V. In Model 1, distance is negatively associated with the likelihood of a paper being referenced by a patent. This is consistent with the contention that firms pay more attention to local knowledge than to knowledge produced further away (Jaffe, Trajtenberg, and Henderson 1993; Alcácer and Chung 2007; Belenzon and Schankerman 2013). In particular, Belenzon and Schankerman (2013) found that the likelihood of a firm citing an academic paper is decreasing in distance, which helps to allay concerns that our dataset of paper twins might exhibit unusual characteristics.

Table V about here

Model 2 of Table V tests whether hubs are really attracting firms' attention to local academic science. It is therefore a more conservative test of the findings obtained in Table III Model 4. In practice, we include an indicator variable measuring whether an author of a paper was located in a hub of relevant

[7] Of course, characteristics of the firm, inventor, or patent itself may influence patent-to-paper citations. In unreported results, we replace simultaneous-discovery/patent fixed effects with fixed effects only for the simultaneous discovery. Although a less strict specification, this allows estimation of patent-level covariates. We fail to establish a correlation between patent-to-paper citations and 1) number of inventors on the patent 2) number of prior patents per inventor 3) number of prior patents awarded to the firm (assignee) 4) whether one of the inventors is in a hub relevant to the focal paper.

commercial R&D. The estimated coefficient on distance diminishes somewhat in magnitude compared to Model 1. Even when controlling for distance, papers located in hubs are 10.1% more likely to be referenced than their twin(s) located outside those hubs, with statistical significance on the estimated coefficient at $p<.023$. This confirms our finding from Figure I and from Model 2 in Table III that the "hub effect" is not simply a manifestation of localized knowledge spillovers, but that hubs are also responsible for spreading the word to distant firms.

One concern is that our results could be driven by particular cities, especially those endowed with specific infrastructure (such as an airport) or those that host large concentrations of either academic scientists or industry inventors. In Model 3 we introduce city fixed effects for the location of the papers' corresponding author. The resulting estimate of the coefficient on hubs has somewhat stronger magnitude and similar statistical significance ($p<.045$). Moreover, the introduction of city fixed effects reduces both the magnitude and statistical significance of the distance covariate to almost nil.

Model 4 adopts an alternative definition of a "hub." Instead of defining a publication as stemming from a hub if any of its authors are within 50 miles from a patenting hotspot, we simply count the (logged) number of patents in the relevant subclasses within 50 miles of the corresponding author's location. Although this measure ignores the fraction of patenting in the subclass and thus does not compare the concentration of relevant R&D activity vs. other locations, it does make use of more information regarding the magnitude of patenting activity. The estimate of the coefficient for the count of patents within 50 miles is positive with statistical significance at the $p<.001$ level.

In Model 5 we repeat the test from Model 7 of Table III: whether academic papers published in a hub are cited not just by firms but also by fellow academics. The analyses in Table V until this point measure references to academic papers from patents assigned to firms in order to measure firms' attention to academic research. If hubs draw attention to local academic papers only from firms, the geography of industrial R&D should have less of an effect on citation rates *from within academia*. For Model 5 we construct dyads of paper twins and university-assigned patents. The estimated coefficient on being in a hub is both diminished in magnitude and imprecisely estimated ($p<.133$) when measuring citations from academic patents to academic papers, as in Table III.

The difference in the magnitude of the hub coefficient after controlling for the type of science in the large sample (+112%) and in the twin sample (+10.1%) is considerable, but a sizeable difference is expected for two main reasons. On one hand, the hub coefficient in the large sample is likely to overestimate the effect of firms' attention. That is because our measures of quality and appliedness in the large sample are imperfect and because this analysis does not allow us to introduce patent-paper-dyad-

level controls such as geographic distance or time. As a result, other factors than firms' attention are likely to inflate the hub coefficient in this analysis. On the other hand, the hub coefficient in Model 2 of table V is likely to underestimate the true effect of firms' selective attention to hub science. That is because the twin papers in our data are by construction well-cited papers, which are important and highly visible. In reality, most academic papers do not emerge as paper twins and do not receive much attention. Presumably, hubs play a larger role in boosting the visibility of more obscure papers than of papers that are already prominent. Thus, our analyses show clearly that firms' attention to science is one of the mechanism explaining the role of hubs in bridging industrial technologies and academic science, but it does not allow us to evaluate precisely the magnitude of this effect.

3.4. Paper Twin Analysis: Mechanisms

The foregoing analysis establishes a connection between the location of academic scientists in hubs and firms' attention as measured by the subsequent likelihood of their articles being cited by industry patents. However, several mechanisms beyond those discussed may be driving our results. Possible mechanisms, including within-twin heterogeneity, formal networks, and informal interactions that arise from co-location of academic and commercial scientists are explored below.

Regarding heterogeneity, although we refer to the simultaneous discoveries as "paper twins" they are not entirely identical. Teams working on the same discovery might use slightly different approaches and describe their findings in different ways, even when the underlying discovery is the same. We assembled an expert panel to evaluate differences between the twin papers, hiring ten postdoctoral researchers in life sciences and applied physics from Imperial College, Cambridge University, University College London, and King's College. We focused their analysis on the 86 papers that belonged to set of twins where one paper stemmed from a hub and the other did not.

The expert-panel investigation was organized in two phases. In phase one, pairs of postdoctoral researchers were asked to jointly evaluate several sets of paper twins and highlight the dimensions along which they differed most. Our panel reported four dimensions along which they could distinguish paper twins: 1) level of detail 2) strength of the claims 3) clarity of exposition 4) clinical relevance. In phase two, we asked the expert panel to evaluate the sets independently and to contrast the papers of the same set along all four dimensions. We then translated the responses into variables, creating dummy variables for each dimension that takes the value 1 if at least one of the postdoctoral researchers believed that one paper was superior to its twin along the specific dimension and 0 otherwise.

Table VI presents results from our expert panel. These complement the findings from Table II. While Table II provides information about the differences in the type of academic science completed inside and

outside of hubs, Table VI provides information about the ways in which academic scientists inside and outside of hubs report scientific findings. The first two columns contain the means of the variables identified by the expert panel for hub and non-hub papers respectively. The third model evaluates whether the mean difference is statistically significant. Heterogeneity is more apparent in some dimensions than in others. For example, our experts were able to tell that one paper was more detailed than the other or that it was more clearly written for about one-third of the pairs. By comparison, they saw differences in the papers' clinical orientation in barely 7% of the pairs. They found that hub-based papers make broader claims than their non-hub twin in 22% of the cases whereas the opposite was true in only 6.7% of the cases ($p<0.042$). On the other hand, we did not find any statistically significant difference in the amount of detail, the clarity, or the clinical orientation of hub- and non-hub-papers.

Table VI about here

The only source of within-twin heterogeneity that clearly differs between papers in hubs vs. those not in hubs is the breadth of claims. However, in Model 1 of Table VII the breadth of claims is not strongly correlated with the likelihood of being cited ($p<0.162$). In fact, the only source of heterogeneity correlated with citation likelihood is clinical orientation (marginal effect = 10.3%; $p<.050$). As noted in Table VI, however, clinical orientation does not differ significantly among hub vs. non-hub papers. Moreover, unreported regressions interacting the hub variable with each measure of heterogeneity fail to achieve statistically significant estimated coefficients on any of the interaction terms. Hence, heterogeneity among papers belonging to the same sets of twins does not seem to explain the hub effect.

Table VII about here

Another possible explanation for the effect might be found in social networks. Singh (2005) shows that citations from one firm's patent to another firm's patent are largely explained by patterns of co-inventorship, but it is unclear whether similar networks would facilitate the flow of knowledge *between* communities such as from academia to industry. We construct networks between academic and industry as the overlap between authors on twin papers and inventors on patents citing them, including second-degree connections (i.e., any of the coauthors of any author on each paper overlapping with any of the co-inventors of any inventor on the patent). Model 2 of Table VII includes our measure of cross-community network overlap. The estimated coefficient is very imprecisely estimated ($p<0.926$). Thus, it does not appear that the hub effect is explained by formal networks.

Finally, a large literature has highlighted the crucial importance of informal interactions for knowledge dissemination in hubs (e.g., Saxenian 1994). In the remaining models of Table VII we examine one observable proxy for informal interactions among commercial and academic scientists as an example

of how these might arise in hubs. Our proxy for informal interactions is the number of conferences held in a specific scientific field. In conducting this analysis, we make two assumptions. First, we assume that academic scientists are, on average, more likely to attend local conferences than distant ones. Not only is the financial cost and inconvenience of travel lower for local scientists; they may more easily "drop in" on a few interesting sessions of a nearby conference whereas travelers from further away must make a larger commitment. Second, we assume that when academic and industrial scientists attend the same conference, knowledge will be more likely to flow between these communities.

Empirical studies have shown that informal interactions such as those taking place at conferences can foster knowledge flow and therefore spur creativity and innovation (Chai, Suchyta, and Freeman 2014; Boudreau et al. 2017). We combined the locations and topics of all Gordon Research Conferences—which are attended by scientists from academia, government, and industry—since 1970 with data from allconferences.com in the same set of scientific fields used to classify organizational prestige. In all, we found 2,383 academic conferences. Importantly, conferences were not exclusively or even primarily located in hubs (781, vs. 1602 in non-hub locations). Popular conference destinations include Bar Harbor, Maine and Santa Fe, New Mexico, which were not hubs for any scientific field among our twin papers. Conferences were geocoded and categorized according to sub-field and then matched to each of the papers if within 50 miles of the principal investigator's location.

In Models 3 and 4 of Table VII we include the number of conferences relevant to the focal paper. The number of nearby conferences in a relevant field is somewhat predictive of the academic paper being cited ($p<0.078$), as seen in model 3, but the hub effect persists. Moreover, in Model 4 a median split of the hub definition on the number of conferences reveals that the hub effect obtains principally among hubs with a higher number of conferences. Relative to non-hub locations, hubs with an above-average number of conferences have a marginal effect on citation of 10.1%, with statistical significance at $p<0.026$. By comparison, the estimate of hubs with a below-average number of conferences is imprecisely estimated ($p<0.613$). In unreported results, the distinction between hubs with more vs. fewer conferences is preserved when splitting by 75th, 90th, or 95th percentile. (Note that all results include a control for the population of the corresponding author's city.) We emphasize that these results should not be interpreted to mean that conferences are the only mechanism by which hubs promote firms' attention to academic science. Rather, they represent an observable event affecting the sort of informal interactions between academic scientists and firms that naturally arise in hubs.

We explore additional mechanisms in Appendix A2. This supplementary analysis shows that hubs attract firms' attention more for academic institutions that have little industry funding or that are relatively unknown in that particular field. We also find that this effect is diminished when the citing firm and the

academic discoverers are either very close or very far from each other. Presumably, firms pay attention to co-located academic discoveries whether or not those are in a hub—but hub location is unlikely to draw the attention of firms located on other continents.

4. Discussion

While innovative firms often exploit academic research to improve their R&D productivity and performance, in practice they rely on only a small fraction of all published articles. In exploring which academic publications firms do and do not draw upon, we posit that hubs act as bridges between academic science and industrial technology. We present two complementary analyses. First, we analyze about 10 million academic publications and their citations in firms' patents to explore the nature of hub- and non-hub publications and their exploitation by firms. Second, we focus on 147 simultaneous discoveries to examine specifically how the fact that an article emerges in a hub affects the attention that it will receive from inventors in firms.

Our results indicate that hubs facilitate the flow of academic science into industrial technologies. In our large-sample analysis, only around 3.8% of all academic publications emerge in hubs, but those publications receive on average 4.6 times more corporate patent citations than non-hub-based papers in the five years following their publication. How can such a large difference be explained?

On the one hand, hub-based publications may simply be more useful from a firm's standpoint. Indeed, we find that academic science from hubs is indeed 28% more applied on average as measured by the keywords associated with those papers. Moreover, hub-based papers appear in journals with a much more applied orientation. We also find large differences in their quality as measured by journal impact factor (+48%) and forward rates of scientific citations (+81%). Still, the apparent effect of hubs on paper citation in firm patents remains large when accounting for those differences. Thus, demand-side factors might also explain in part the relatively high exploitation of academic science from hubs.

We posit that firms pay disproportionate attention to academic knowledge when it emerges in hubs. They do so not only because they use hub-origin as a cue that a publication deserves their attention, but also because they are likely to be more exposed to academic discoveries made in hubs. In our large-sample analysis, several correlations are supportive of this interpretation. First, firms' reliance on hubs is much stronger when a discovery is first published and thus uncertainty regarding its usefulness is high, diminishing as the science ages. Second, specialized hubs have a stronger effect than diversified hubs. Third, university-industry collaborations attract the attention of other firms and therefore substitute for the effect of hub location. Fourth, firms that are younger, more innovative, and more science-oriented disproportionately draw from hubs.

Although these findings suggest that firms allocate their attention toward hub-based science, unobserved heterogeneity among papers makes it difficult to make a strong claim. We therefore explore the role of firms' attention by focusing on a set of 697 corporate patents and their citations to 147 simultaneous discoveries embodied in 316 paper twins. We find that firms are significantly more likely to cite an academic article in their patents if it has an author in a hub of industrial R&D than the twin from outside a hub. We then explore mechanisms that may explain this relatively high demand for hub-based academic science. Clearly, firms might allocate more attention to academic discoveries produced in hubs if they expect that those discoveries will be on average more useful for them. However, other mechanisms are also possible.

First, the differential citation between hub- and non-hub papers may be driven by geographic distance since hub-produced academic papers are collocated with a larger number of potential citing inventors. We control for geographic distance in all regressions using the paper twins and find that the apparent effect of hubs is particularly strong when firms are not collocated with the academic discoverers. Second, the effect could in principle be driven by city-invariant characteristics (e.g., infrastructure such as airports) but we find that it remains robust to the addition of city fixed effects. Third, there may be subtle differences between hub- and non-hub papers which drive our result. We hired an expert panel to evaluate 86 papers that belonged to sets of twins in which one had emerged from a hub and the other did not. The subtle within-twin differences that were identified did not explain the gap in patent citations to hub- and non-hub papers. Fourth, formal networks between citing corporate inventors and cited academic authors may explain our result. We examined the overlap between authors on twin papers and inventors on patents citing them, including second-degree connections, but those did not predict citation. Finally, we studied whether the type of informal interactions that take place in hubs might explain our results. In line with this mechanism, we found the effect of location to be particularly strong for those that host a greater number of conferences where firms and academics can meet. Thus, our results suggest that firms display relatively high demand for hub-based science both because they expect it to be more useful and because informal interactions in hubs mean that they are more likely to be exposed to those papers.

Our dual empirical exercise thus provides two distinct but complementary perspectives on the role of hubs as cross-institutional bridges. However, it is not without limitations and therefore presents important avenues for future research. For example, as described in the twin analysis results section, we are unable to estimate precisely the magnitude of the effect of hubs on firms' attention for non-twin papers. We are also unable to assess the extent to which firms' disproportionate attention to hub science is driven by their active monitoring of hubs as opposed to their passive reaction to their differential exposure to hub-based and non-hub based science. (Note that the result regarding hubs and conferences is compatible with either

of these interpretations.) Nor can we fully assess *why* hubs produce academic research that appears to be more useful for technological innovation. Perhaps most importantly, we do not assess whether focusing on science from hubs improves or limits firms' innovation prospects - potentially a focus for future research.

Our study makes two main contributions to theory. First, the "scientification" of commercial technology (Henrekson and Rosenberg 2001) raises important questions about firms' ability to exploit academic findings. While prior research has explored the flow of university science to firms, it has mostly emphasized science's tacit nature or stickiness (von Hippel 1994; Zucker, Darby, and Armstrong 2002; Agrawal 2006); few studies have considered how firms allocate their limited attention to a growing academic literature. We contribute to this line of inquiry by describing the crucial role of hubs in the institutional hand-off, highlighting how the interface between academia and industry is embedded in a particular geography. Hubs of industrial R&D not only concentrate high-quality and commercially relevant academic science, they also attract attention from a large number of innovative firms.

Our second contribution is to clarify the nature of knowledge recombination in hubs. Marshall (1890:7) famously observed that "great are the advantages people following the same skilled trade get from the near neighbourhood…if one man starts a new idea, it is taken up by others," which has been borne out by findings that knowledge flows most smoothly from academia to industry over short distances (Jaffe, Tratjenberg, & Henderson 1993; Belenzon & Schankerman, 2013). Our findings complement this stream of literature by highlighting that hubs facilitate the flow of knowledge not only within their boundaries but beyond. Hubs foster distant knowledge flows by attracting the attention of firms located hundreds of miles away. This is important because knowledge flows between academia and industry—at least as measured by patent-to-paper citations—are primarily distant, not local.

Our study also has implications for firm strategy. Our results indicate that firms seeking to take advantage of academic science face a tradeoff. On one hand, by focusing their attention on hubs, innovative firms have access to high quality and relevant science while keeping search costs relatively low. However, those same discoveries are likely to be visible to other firms, including competitors, and they are therefore unlikely to be a source of competitive advantage. On the other hand, firms choosing to search "beyond the lamppost" are likely to face much higher search costs because academic knowledge outside of hubs is vast, spread out, and generally of a more fundamental nature and lower quality. However, discoveries outside of hubs receive also less attention. Hence, there are presumably more hidden gems—in the form of undervalued academic discoveries such as Goldwasser's purified erythropoietin—outside of hubs than inside them.

Finally, we provide data artifacts to facilitate future research. Our hubs for patent subclasses are posted publicly for others to download, and we also make available a new measure of journal relevance, the Journal Commercial Impact Factor (JCIF), created using an algorithm similar to JIF but counting citations from industry patents instead of from papers. (The patent-to-WoS citations we use from Fleming et al. (2017) are not publicly available, but we refer interested researchers to Marx and Fuegi (2019) for a downloadable set of citations from USPTO patents 1947-2018 to the Microsoft Academic Graph 1800-2018.) Our hope is that these new definitions and constructs, along with the underlying data, will enable others to extend this line of inquiry.

Academic science constitutes a formidable resource for innovative firms (Fleming and Sorenson 2004; Li, Azoulay, and Sampat 2017; Arora, Belenzon, and Patacconi 2018), but its exploitation can be challenging (e.g., Murray 2010; Bikard 2018). By clarifying the intricate relationship between academic science and firms' innovation, we hope that our work—as well as future research—on this topic will make it easier for individuals, firms, and nations to find in the academic literature a source of competitive advantage.

References

Agrawal, Ajay. 2006. "Engaging the Inventor: Exploring Licensing Strategies for University Inventions and the Role of Latent Knowledge." *Strategic Management Journal* 27 (1): 63–79.

Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. 2016. "Understanding the Changing Structure of Scientific Inquiry." *American Economic Journal: Applied Economics* 8 (1): 100–128.

Agrawal, Ajay, and Rebecca M. Henderson. 2002. "Putting Patents in Context: Exploring Knowledge Transfer from MIT." *Management Science* 48 (1): 44-60.

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2017. "The Dual Frontier: Patented Inventions and Prior Scientific Advance." *Science* 357 (6351): 583–87.

Alcácer, Juan, and Wilbur Chung. 2007. "Location Strategies and Knowledge Spillovers." *Management Science* 53 (5): 760–76.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi. 2018. "The Decline of Science in Corporate R&D." *Strategic Management Journal* 39 (1): 3–32.

Arora, Ashish, and Alfonso Gambardella. 1994. "The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labour." *Research Policy* 23 (5): 523–532.

Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2013. "Matthew: Effect or Fable?" *Management Science* 60 (1): 92–109.

Begley, C. Glenn, and Lee M. Ellis. 2012. "Drug Development: Raise Standards for Preclinical Cancer Research." *Nature* 483 (7391): 531–33.

Belenzon, Sharon, and Mark Schankerman. 2013. "Spreading the Word: Geography, Policy, and Knowledge Spillovers." *Review of Economics and Statistics* 95 (3): 884–903.

Bena, Jan, and Kai Li. 2014. "Corporate Innovations and Mergers and Acquisitions." *The Journal of Finance* 69 (5): 1923–60.

Bikard, Michaël. 2012. "Simultaneous Discoveries as a Research Tool: Method and Promise." *MIT Sloan Working Paper*.

———. 2018. “Made in Academia: The Effect of Institutional Origin on Inventors’ Attention to Science.” *Organization Science* 29 (5): 818–36.
Bikard, Michaël, Keyvan Vakili, and Florenta Teodoridis. 2019. “When Collaboration Bridges Institutions: The Impact of University–Industry Collaboration on Academic Productivity.” *Organization Science*, 30 (2):426-45.
Boudreau, Kevin J., Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R. Lakhani. 2017. “A Field Experiment on Search Costs and the Formation of Scientific Collaborations.” *The Review of Economics and Statistics* 99 (4): 565-76.
Bush, Vannevar. 1945. “Science: The Endless Frontier.” *Transactions of the Kansas Academy of Science (1903-)* 48 (3): 231–64.
Callaert, Julie, Maikel Pellens, and Bart Van Looy. 2014. “Sources of Inspiration? Making Sense of Scientific References in Patents.” *Scientometrics* 98 (3): 1617–29.
Cassiman, Bruno, Reinhilde Veugelers, and Pluvia Zuniga. 2008. “In Search of Performance Effects of (in)Direct Industry Science Links.” *Industrial and Corporate Change* 17 (4): 611–46.
Chai, Sen, and Anoop Menon. 2019. “Breakthrough Recognition: Bias against Novelty and Competition for Attention.” *Research Policy* 48 (3): 733–47.
Chai, Sen, and Willy Shih. 2016. “Bridging Science and Technology through Academic–Industry Partnerships.” *Research Policy* 45 (1): 148–58.
Chai, Sen, Marissa Suchyta, and Richard B. Freeman. 2014. “Conferences as a Vehicle for Collaborative Discovery and Innovation.” *Academy of Management Proceedings* 2014 (1): 16246.
Cohen, Wesley, Richard Nelson, and John Walsh. 2002. “Links and Impacts: The Influence of Public Research on Industrial R&D.” *Management Science* 48 (1): 1–23.
Cozzens, Susan E. 1989. *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*. State University of New York Press.
D’Este, Pablo, and Markus Perkmann. 2011. “Why Do Academics Engage with Industry? The Entrepreneurial University and Individual Motivations.” *The Journal of Technology Transfer* 36 (3): 316–39.
Drivas, Kyriakos, Zhen Lei, and Brian D. Wright. 2017. “Academic Patent Licenses: Roadblocks or Signposts for Nonlicensee Cumulative Innovation?” *Journal of Economic Behavior & Organization* 137 (May): 282–303.
Evans, James A. 2010. “Industry Induces Academic Science to Know Less about More.” *The American Journal of Sociology* 116 (2): 389–452.
Fleming, Lee, Guan-Cheng Li, Matt Marx, Dennis Yao, and Hillary Greene. 2017. “U.S. Innovation Relies Increasingly Upon Federal Research Investment.”
Fleming, Lee, and Olav Sorenson. 2004. “Science as a Map in Technological Search.” *Strategic Management Journal* 25: 909–28.
Freedman, Leonard P., Iain M. Cockburn, and Timothy S. Simcoe. 2015. “The Economics of Reproducibility in Preclinical Research.” *PLOS Biology* 13 (6): e1002165.
Furman, Jeffrey, and Megan J. MacGarvie. 2007. “Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry.” *Journal of Economic Behavior & Organization* 63 (4): 756–76.
Gittelman, Michelle. 2007. “Does Geography Matter for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology.” *Organization Science* 18 (4): 724–41.
Gittelman, Michelle, and Bruce Kogut. 2003. “Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns.” *Management Science* 49 (4): 366–82.
Goldwasser, Eugene. 2011. *A Bloody Long Journey*. Xlibris Corporation.
Halim, Nadia. 2000. “Bridging Apoptotic Signaling Gaps.” *The Scientist*, August 21, 2000.
Hartshorne, Joshua K., and Adena Schachner. 2012. “Tracking Replicability as a Method of Post-Publication Open Evaluation.” *Frontiers in Computational Neuroscience* 6 (March).

Henrekson, Magnus, and Nathan Rosenberg. 2001. "Designing Efficient Institutions for Science-Based Entrepreneurship: Lesson from the US and Sweden." *The Journal of Technology Transfer* 26 (3): 207–31.
Hippel, Eric von. 1994. "'Sticky Information' and the Locus of Problem Solving: Implications for Innovation." *Management Science* 40 (4): 429–39.
Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science." *The Quarterly Journal of Economics* 133 (2): 927–91.
Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3): 577–98.
Jinha, Arif E. 2010. "Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence." *Learned Publishing* 23 (3): 258–63.
Jones, Benjamin. 2009. "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76 (1): 283–317.
Leahey, Erin. 2007. "Not by Productivity Alone: How Visibility and Specialization Contribute to Academic Earnings." *American Sociological Review* 72 (4): 533–561.
Lemley, Mark A., and Bhaven Sampat. 2011. "Examiner Characteristics and Patent Office Outcomes." *Review of Economics and Statistics* 94 (3): 817–27.
Li, Danielle, Pierre Azoulay, and Bhaven N. Sampat. 2017. "The Applied Value of Public Investments in Biomedical Research." *Science* 356 (6333): 78–81.
Marx, Matt, and Aaron Fuegi. 2019. "Reliance on Science in Patenting." SSRN Scholarly Paper ID 3331686. Rochester, NY: Social Science Research Network.
Merton, Robert K. 1961. "Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science." *Proceedings of the American Philosophical Society* 105 (5): 470–86.
———. 1968. "The Matthew Effect in Science The Reward and Communication Systems of Science Are Considered." *Science* 159 (3810): 56.
Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.
Murray, Fiona. 2002. "Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering." *Research Policy* 31 (8–9): 1389–1403.
———. 2010. "The Oncomouse That Roared: Hybrid Exchange Strategies as a Source of Distinction at the Boundary of Overlapping Institutions1." *The American Journal of Sociology* 116 (September): 341–88.
Nelson, Richard. 1982. "The Role of Knowledge in R&D Efficiency." *The Quarterly Journal of Economics*, The Quarterly Journal of Economics, 97 (3): 453–70.
Ocasio, William. 1997. "Towards an Attention-Based View of the Firm." *Strategic Management Journal* 18: 187–206.
———. 2010. "Attention to Attention." *Organization Science* 22 (5): 1286–96.
Osherovich, Lev. 2011. "Hedging against Academic Risk." *SciBX: Science-Business EXchange* 4 (15).
Podolny, J. M., T. E. Stuart, and M. T. Hannan. 1996. "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991." *American Journal of Sociology*, 659–89.
Polidoro, Francisco. 2013. "The Competitive Implications of Certifications: The Effects of Scientific and Regulatory Certifications on Entries into New Technical Fields." *Academy of Management Journal* 56 (2): 597–627.
Reschke, Brian P., Pierre Azoulay, and Toby E. Stuart. 2018. "Status Spillovers: The Effect of Status-Conferring Prizes on the Allocation of Attention." *Administrative Science Quarterly* 63 (4): 819–47.
Roach, Michael, and Wesley M. Cohen. 2013. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science* 59 (2): 504–25.
Rosenberg, Nathan, and Richard Nelson. 1994. "American Universities and Technical Advance in

Industry." *Research Policy* 23 (3): 323–48.
Rosenblatt, Michael. 2016. "An Incentive-Based Approach for Improving Data Reproducibility." *Science Translational Medicine* 8 (336): 336ed5-336ed5.
Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press.
Simcoe, Timothy S., and Dave M. Waguespack. 2011. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* 57 (2): 274–90.
Simon, Herbert. 1947. *Administrative Behavior*. Free Press.
Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–70.
Singh, Jasjit, and Matt Marx. 2013. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." *Management Science* 59 (9): 2056–78.
Sohn, Eunhee. 2014. "The Endogeneity of Academic Science to Local Industrial R&D." *Academy of Management Proceedings* 2014 (1): 11413.
Stern, Scott. 2004. "Do Scientists Pay to Be Scientists?" *Management Science* 50 (6): 835–53.
Stokes, Donald E. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press.
Teodoridis, Florenta, Michaël Bikard, and Keyvan Vakili. 2018. "Creativity at the Knowledge Frontier: The Impact of Specialization in Fast- and Slow-Paced Domains." *Administrative Science Quarterly*, July, 0001839218793384.
Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong. 2002. "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology." *Management Science* 48 (1): 138–53.
Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *The American Economic Review* 88 (1): 290–306.
Zuckerman, E. W., and T. Y. Kim. 2003. "The Critical Trade-off: Identity Assignment and Box-Office Success in the Feature Film Industry." *Industrial and Corporate Change* 12 (1): 27–67.

Table I: Large-sample: descriptive statistics and correlations

Variable	N	Mean	Std. Dev.	Min	Max	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)
# of cites from patents (5yrs)	10,169,440	0.06	1.1	0	769	1.00										
Year published	10,169,440	2005.3	9.4	1900	2018	0.00	1.00									
Paper in a hub	10,169,440	0.038	0.190	0	1	0.03	0.01	1.00								
Paper in US	10,169,440	0.322	0.467	0	1	0.01	-0.15	0.18	1.00							
# cites from papers (5yrs)	10,169,440	15.110	43.440	0	38042	0.05	-0.02	0.06	0.10	1.00						
Journal impact factor	10,169,442	1.819	2.112	0	351.667	0.03	0.04	0.08	0.12	0.39	1.00					
Journal commercial impact factor	10,169,442	0.013	0.039	0	4.53571	0.07	-0.01	0.08	0.08	0.22	0.41	1.00				
Appliedness of paper	9,820,968	0.026	0.026	0	2.89	0.09	0.01	0.06	0.01	0.18	0.28	0.38	1.00			
Institution's prestige in paper field	10,169,442	944.647	2230.406	0	30062.00	0.02	-0.10	0.09	0.26	0.16	0.23	0.13	0.15	1.00		
Authors' prestige in paper field	10,169,442	1.71	3.42	0	314	0.02	-0.28	0.03	0.03	0.24	0.13	0.12	0.12	0.22	1.00	
Academic collaboration w/industry	10,169,442	0.089	0.285	0	1.00	0.14	0.09	0.01	-0.02	0.03	0.03	0.03	0.05	-0.01	-0.01	1.00

Table II: Large-sample: difference-of-means tests

	papers in hubs	*papers not in hubs*	*std. err*
# cites from patents (5yrs)	0.223	0.049	0.0001
Year published	2005.7	2005.3	0.0154
Paper in US	0.747	0.305	0.0008
# cites from papers (5yrs)	26.575	14.663	0.0018
Journal impact factor	2.640	1.787	0.0035
Journal commercial impact factor	0.028	0.013	0.0001
Appliedness of paper	0.033	0.025	0.0000
Institution's prestige in paper field	1795.237	911.496	0.0036
Authors' prestige in paper field	2.065	1.695	0.0011
Academic collaboration w/industry	0.103	0.088	0.0005

Notes: Data are from academic articles reported by the Web of Science for which hubs could be determined by assigning at least one USPTO patent subclass to the paper. Patent subclasses were assigned to papers by observing either 1) patent citations to the paper, or 2) when the paper was not cited by any patent, by examining patent citations to all the papers sharing the same keywords as determined by PubMed and/or the Microsoft Academic Graph. Appliedness is calculated only when PubMed or Microsoft Academic Graph keywords are available, resulting in a smaller number of observations for this variable.

Table III: Large sample: Patent citations to hub and non-hub publications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DV:citations to academic papers from	*industry patents*						*university patents*
Paper in a hub	1.230***	1.252***	0.762***	0.752***	0.812***		0.462***
	(0.0266)	(0.0324)	(0.0271)	(0.0263)	(0.0351)		(0.0467)
Paper in US	0.379***	0.417***	0.113***	0.0616**	0.0662**	0.0712**	0.627***
	(0.0248)	(0.0297)	(0.0317)	(0.0312)	(0.0312)	(0.0308)	(0.0187)
Ln # cites from papers (5yrs)			0.715***	0.699***	0.688***	0.698***	0.779***
			(0.0229)	(0.0225)	(0.0218)	(0.0222)	(0.0172)
Journal impact factor			-0.0117**	-0.0171***	-0.0163***	-0.0168***	0.0100**
			(0.0059)	(0.0056)	(0.0056)	(0.0056)	(0.0043)
Journal commercial impact factor			1.347***	1.372***	1.374***	1.372***	1.394***
			(0.165)	(0.163)	(0.161)	(0.163)	(0.134)
Ln appliedness of paper			8.424***	8.491***	8.462***	8.518***	8.118***
			(0.737)	(0.731)	(0.696)	(0.725)	(0.650)
Ln institution's prestige in paper field				0.0524***	0.0542***	0.0525***	0.0844***
				(0.0057)	(0.0056)	(0.0056)	(0.0071)
Ln authors' prestige in paper field				0.110***	0.112***	0.109***	0.124***
				(0.0167)	(0.0157)	(0.0167)	(0.0164)
Academic collaboration w/industry					0.536***		
					(0.0246)		
Industry author * hub					-0.442***		
					(0.145)		
Paper in a diverse hub						0.673***	
						(0.0249)	
Paper in a specialized hub						1.319***	
						(0.0685)	
Distance restriction on patent cites?	no	>50mi	no	no	no	no	no
N	10,156,451	10,156,451	9,808,083	9,808,083	9,808,083	9,808,083	9,805,612

Notes: Each observation is an academic paper reported by the Web of Science for which hubs could be determined. All models are estimated using Poisson with year-level fixed effects. The dependent variable is the number of citations from firm-assigned patents to each of these academic papers within 5 years of publication. *** p<0.01; ** p<0.05; * p<.0.1.

Table IV: Large sample: Patent to paper citations for different types of firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DV =	*% of papers cited by firm's patents that are in hubs*									
Assignee age (years since first patent)	-0.0321***	-0.0176***							-0.0432***	-0.0296***
	(0.0021)	(0.0029)							(0.0023)	(0.0033)
Assignee size (# of patents)			0.0039***	0.0032***					0.0123***	0.00857***
			(0.0006)	(0.0007)					(0.0007)	(0.0009)
Assignee patent quality (avg. fwd cites)					0.0152***	0.0101***			0.0157***	0.0102***
					(0.0011)	(0.0014)			(0.0011)	(0.0015)
Assignee reliance on science							0.0129***	0.0080***	0.0181***	0.0104***
							(0.0013)	(0.0017)	(0.0014)	(0.0019)
Constant	0.181***	0.143***	0.0746***	0.0821***	0.0661***	0.0776***	0.0750***	0.0837***	0.152***	0.137***
	(0.0064)	(0.0087)	(0.0018)	(0.0021)	(0.0016)	(0.0020)	(0.0013)	(0.0017)	(0.0070)	(0.0098)
firm's most frequent patent-subclass FE	no	yes	no	yes	no	yes	no	yes	no	yes
N	36,769	28,976	36,769	28,976	36,769	28,976	36,769	28,976	36,769	28,976
R-squared	0.006	0.265	0.001	0.265	0.005	0.266	0.003	0.265	0.021	0.270

Notes: Each observation is an assignee citing at least one publications in its patent for which hubs could be determined. The number is observations is lower in even-numbered models as firms for whom all of their patents are in a single patent subclass are dropped. All models are estimated using OLS. *** p<0.01; ** p<0.05; * p<.0.01.

Table V: The impact of the location of academic institutions on citation of twin papers by patents

	(1)	(2)	(3)	(4)	(5)
	DV: cites from industry patents				*DV: cites from academic patents*
Paper author(s) in a hub		0.717*	1.024*		0.496
		(0.316)	(0.514)		(0.330)
Ln # patents within 50 miles of corresponding author				0.929***	
				(0.183)	
Ln distance between paper and patent	-0.244***	-0.171**	-0.108	-0.157**	0.0197
	(0.0626)	(0.0559)	(0.0916)	(0.0563)	(0.0751)
N	1,671	1,671	1,671	1,671	1,088
# twin articles	316	316	316	316	381
paper, author, institution controls	no	yes	yes	yes	yes
City fixed effects	no	no	yes	no	no
Simultaneous-discovery/patent FE	yes	yes	yes	yes	yes

Notes: Observations are academic-paper/industry-patent dyads. All models are estimated using conditional logit and include simultaneous-discovery/patent fixed effects. Controls include those for the paper (U.S.-based, journal impact factor, paper was patented), corresponding author (stock of patents and papers, population of city), and institution (stock of patents and papers) characteristics as well as characteristics of the paper-patent dyad (publication lag). Standard errors are clustered at the level of the simultaneous discovery; *** p<0.001; ** p<0.01; * p<.0.05; + p<0.1.

Table VI: Difference of means tests for within-twin heterogeneity

	Papers in hubs	Papers not in hubs	*p* <
Paper is more detailed	0.390	0.311	0.447
Paper has broader claims	0.220	0.067	0.042
Paper is more clearly written	0.414	0.266	0.151
Paper is more clinically oriented	0.073	0.067	0.907

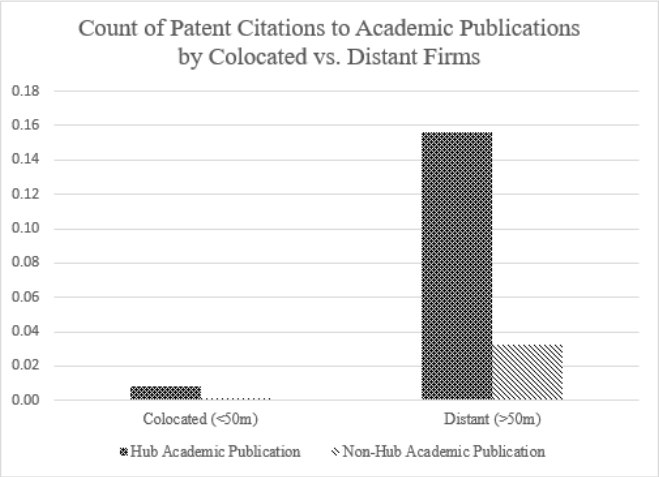
Notes: Results of expert-panel evaluation of 86 "twin" papers by ten postdoctoral researchers in the life sciences and applied physics. Postdocs identified the four criteria in the first column and classified twin papers accordingly.

Table VII: Mechanisms

	(1)	(2)	(3)	(4)
	DV: citations from industry patents			
Paper author(s) in a hub	1.575***	0.716*	0.641*	
	(0.467)	(0.312)	(0.307)	
Paper published before its twin(s)	-0.285			
	(0.609)			
Paper more detailed than its twin(s)	0.382			
	(0.604)			
Paper has more claims than its twin(s)	-0.849			
	(0.608)			
Paper written more clearly than its twin(s)	0.523			
	(0.416)			
Paper more clinical than its twin(s)	1.346+			
	(0.688)			
Coauthor of a paper author is an inventor on paten		0.0363		
		(0.391)		
Ln # conferences in this field held within 50 miles			0.150+	
of corresponding author			(0.0851)	
Author(s) in hub * >median # of conferences				0.733*
				(0.330)
Author(s) in hub * <median # of conferences				0.460
				(0.908)
Ln distance between paper and patent	-0.208	-0.171**	-0.137**	-0.170**
	(0.163)	(0.0566)	(0.0509)	(0.0565)
N	552	1,671	1,671	1,671
# twin articles	78	316	316	316
paper, author, institution controls	yes	yes	yes	yes
Simultaneous-discovery/patent FE	yes	yes	yes	yes

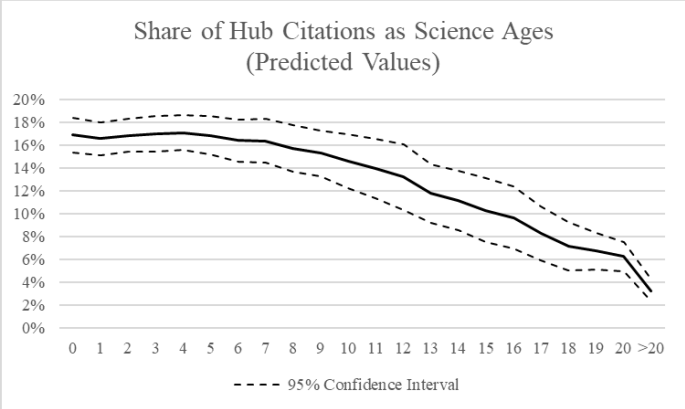
Notes: All models include controls for characteristics of the paper (U.S.-based, journal impact factor, discovery was patented), author (stock of patents and papers, city population), institution (stock of patents and papers), and paper-patent dyad (publication lag; spatial distance). Model 1 contains fewer observations as its analysis is limited to the twin papers evaluated for heterogeneity by the expert panel. Standard errors are clustered throughout at the level of the simultaneous discovery; *** p<0.001; ** p<0.01; * p<.0.05; + p<0.1.

Figure I: Local versus distant knowledge flow



Notes: Data are from academic articles reported by the Web of Science for which hubs could be determined by assigning at least one US patent subclass to the paper.

Figure II: Firms turn to hubs for more recent science



Notes: Estimated using linear probability model predicting hub origin as a function of dichotomized citation age with year fixed effects and robust standard errors clustered at the level of the year in a population of 2,674,170 patent-to-paper citations. Citation Age = Patent Application Year – Paper Publication Year.

Figure III: Example of a set of "paper twins" reporting a simultaneous discovery

Cell, Vol. 94, 481–490, August 21, 1998, Copyright ©1998 by Cell Press

Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors

Xu Luo,[1] Imawati Budihardjo,[1] Hua Zou, Clive Slaughter, and Xiaodong Wang*
Howard Hughes Medical Institute
and Department of Biochemistry
University of Texas Southwestern Medical Center at Dallas
Dallas, Texas 75235

Summary

We report here the purification of a cytosolic protein that induces cytochrome c release from mitochondria in response to caspase-8, the apical caspase activated by cell surface death receptors such as Fas and TNF. Peptide mass fingerprinting identified this protein as Bid, a BH3 domain–containing protein known to interact with both Bcl2 and Bax. Caspase-8 cleaves Bid, and the COOH-terminal part translocates to mitochondria where it triggers cytochrome c release. Immunodepletion of Bid from cell extracts eliminated the cytochrome c releasing activity. The cytochrome c releasing activity of Bid was antagonized by Bcl2. A mutation at the BH3 domain diminished its cytochrome c releasing activity. Bid, therefore, relays an apoptotic signal from the cell surface to mitochondria.

Cell, Vol. 94, 491–501, August 21, 1998, Copyright ©1998 by Cell Press

Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis

Honglin Li, Hong Zhu, Chi-jie Xu, and Junying Yuan*
Department of Cell Biology
Harvard Medical School
Boston, Massachusetts 02115

Summary

We report here that BID, a BH3 domain–containing proapoptotic Bcl2 family member, is a specific proximal substrate of Casp8 in the Fas apoptotic signaling pathway. While full-length BID is localized in cytosol, truncated BID (tBID) translocates to mitochondria and thus transduces apoptotic signals from cytoplasmic membrane to mitochondria. tBID induces first the clustering of mitochondria around the nuclei and release of cytochrome c independent of caspase activity, and then the loss of mitochondrial membrane potential, cell shrinkage, and nuclear condensation in a caspase-dependent fashion. Coexpression of Bcl2 inhibits all the apoptotic changes induced by tBID. Our results indicate that BID is a mediator of mitochondrial damage induced by Casp8.