Paper to be presented at the DRUID 2012

on

June 19 to June 21

at

CBS, Copenhagen, Denmark,

Is Academic Science Trapped Inside The Ivory Tower? Universities and Science-Based Cumulative Invention

Michaël Bikard
MIT
Sloan School of Management
mbikard@mit.edu

Abstract
We examine two views of the emergence of science-based technologies. In one, the academic environment encourages cumulative inventions through its culture of sharing and openness. In the other, firms offer a more fertile ground for science-based technologies because of their distinctive mission, larger resources and deeper connections with inventors. Analysis of cumulative invention based on 39 simultaneous discoveries between academia and industry involving 90 teams and cited in 533 patents shows that the most prolific source of science-based inventions are firms. Industry discoverers generate over 3 times as many cumulative patents as their co-discoverers in academia. Moreover, non-discoverer inventors draw scientific knowledge from firms at a 10-20% higher rate than they do from academia.

Jelcodes:O32,-

Is Academic Science Trapped Inside The Ivory Tower? Universities and Science-Based Cumulative Invention

Michaël Bikard
PhD Candidate
MIT Sloan School of Management
mbikard@mit.edu

April 2012

ABSTRACT

We examine two views of the emergence of science-based technologies. In one, the academic environment encourages cumulative inventions through its culture of sharing and openness. In the other, firms offer a more fertile ground for science-based technologies because of their distinctive mission, larger resources and deeper connections with inventors. Analysis of cumulative invention based on 39 simultaneous discoveries between academia and industry involving 90 teams and cited in 533 patents shows that the most prolific source of science-based inventions are firms. Industry discoverers generate over 3 times as many cumulative patents as their co-discoverers in academia. Moreover, non-discoverer inventors draw scientific knowledge from firms at a 10-20% higher rate than they do from academia.

Keywords: Corporate R&D; Academic science; Knowledge diffusion; Simultaneous discoveries; Patents

1. INTRODUCTION

A large and growing literature has argued that scientific research has a positive impact on technological innovation in both firms (Cohen and Levinthal 1990) and universities (Henderson, Jaffe, and Trajtenberg 1998). However much less is known of the conditions under which the scientific knowledge produced in firms and universities is recombined into novel technologies. Historical examples reveal that while this translation can be very rapid in some circumstances, scientific knowledge can also remain unexploited for years, even decades, before an inventor finally uses it—if at all[1] (Rosenberg 1994; Mokyr 2002). In this paper, we try to fill this gap by examining the factors that lead to the translation of scientific discoveries into novel technologies, a process we term "technology spawning." We examine two views of this process.

In one view, the academic environment allows for a large diffusion of new knowledge and therefore fosters cumulative invention. The diffusion of knowledge from academic labs is not only a consequence of the educational mission of universities; it also results from the academic norms of sharing and openness that sharply contrast with the use of secrecy and exclusion of commercial science (Dasgupta and David 1994; Murray 2010). In this view, scientific research performed in academia should therefore be recombined into a larger number of technologies than if it were produced in industry (David and Dasgupta 1987; Zucker, Darby, and Brewer 1998; Henderson, Jaffe, and Trajtenberg 1998; Adams 2002; Owen-Smith and Powell 2004; Furman et al. 2005). We refer to this description of technology spawning as the "Entrepreneurial University view."

Another view of the technology spawning process argues instead that the corporate environment leads to higher rates of cumulative invention. Unlike academic scientists working in the ivory tower, industry scientists have clear incentives, better knowledge, and more resources to develop new technologies (Zucker, Darby, and Armstrong 2002; Aghion, Dewatripont, and Stein 2008). Knowledge often flows between competing firms (Saxenian 1994) and industrial R&D scientists primarily reuse knowledge from other firms rather than knowledge from academia (Cohen, Nelson, and Walsh 2002). Since scientific and technological networks tend to be distinct (Murray 2002), new knowledge emerging in industry might be accessible to a larger number of inventors. We refer to this as the "Ivory Tower view" of technology spawning.

---

[1] For instance, the Hellenistic civilization produced Ptolemaic Astronomy but never used it for navigation, they understood optics but did not translate that knowledge into making binocular or glasses (Mokyr 2002, 262)

This paper is an empirical exploration of these two views of the technology spawning process. While we frame the paper as a "horse race" between the two hypotheses, our analysis clearly shows that both types of environments contribute to the creation of new technologies. Academic science emerges as an important source of science-based invention, as does industrial R&D. While both types of spawning emerge in practice, we hope to shed light on the respective impact of the academic and corporate research environments on cumulative invention.

The empirical challenge in testing these hypotheses is straightforward. Observed rates of cumulative inventions might result from the environment in which a given discovery was made but it might also be a consequence of the nature and promises of that discovery. For instance, evidence that scientific publications authored by firms receive more patent citations than academic papers might indicate that private firms constitute a more favorable environment for cumulative invention. However, it might also reflect the fact that discoveries made in universities tend to be more basic than discoveries made in firms. The implications are vastly different depending on the mechanism. In order to compare the two views of technology spawning, it is crucial to account for the technological potential of the observed scientific discoveries. This paper reports a novel empirical strategy to tackle this question.

In the winter of 1999, two teams of scientists simultaneously discovered VR1 (vanilloid receptor-1), the receptor for the pain caused by excessive heat or capsaicin, the pungent component of chili peppers. The first team, led by Dr John B Davis, sent its results to *Nature* on December 20, 1999 and the paper was published on May 11, 2000. The second team, led by Prof David Julius, sent its results to *Science* on January 18, 2000 and the paper was published on April 14, 2000. The new knowledge had important implications for the development of pain therapeutics. Yet, both discoveries were made in very different environments. Julius is an academic based at UC San Francisco. In contrast, Davis is an industrial scientist working at Smithkline Beecham in Harlow, UK.

Simultaneous discoveries are a fascinating and relatively frequent phenomenon which has historically received a lot of attention (Merton 1961). When the discoverers submit their findings for publication at almost the same time, two or more papers disclosing the same discovery can be accepted, thus leading to the publication of "paper-twins." Paper-twins are scientific articles that disclose the same underlying piece of knowledge. They are thus more than complementary discoveries or discoveries belonging to the same scientific specialty. Rather, by embedding the same piece of knowledge that emerged in two distinct environments, paper-twins are a natural consequence of the duplication of effort in science, and a potentially rich setting to study the determinants of cumulative invention.

The “experiment” afforded by the observation of simultaneous discoveries occurring in an academic and an industrial environment will allow for a set of precise tests motivated by the two competing views on technology spawning. The citations of each twin paper in the patent literature provide indeed a convenient (though noisy) measure of cumulative invention. First, we study the extent to which the discoverers in each environment will invent based on their scientific discovery. Second, we explore the use of scientific knowledge from academic and industrial laboratories by inventors who did not take part in the discovery. Because it enables the observation of the *non-occurrence of cumulative inventions that could have occurred*, paper-twins are a particularly suitable setting to investigate the impact of the research environment on cumulative invention.

The analysis centers on 39 simultaneous discoveries made by 90 teams and that involved at least one team from academia and another team from a firm. These 90 papers are cited in 533 patents, therefore allowing for a quantitative examination of the two views on technology spawning. The chosen setting is narrower in scope than the massive data-based efforts analyzing tens of thousands of academic publications and patents (Narin, Hamilton, and Olivastro 1997; Henderson, Jaffe, and Trajtenberg 1998) or large-scale survey data (Klevorick et al. 1995; Cohen, Nelson, and Walsh 2002) but larger than qualitative small-scale investigations (Colyvas et al. 2002). The analysis establishes two core findings. First, we find that discoverers working in industry generate more than three times the number of patents of their co-discoverers in academia. Second, the data indicates that non-discoverer inventors tend to draw their scientific knowledge from firms rather than universities. A discovery made in a firm is 10-20% more likely to be cited in follow-on patents than its academic twin. Taken together, these results provide strong support for the view that the corporate environment leads to the higher rates of cumulative invention.

2. SCIENCE-BASED CUMULATIVE INVENTION

Science is broadly considered a key input into new technologies (Rosenberg 1994; Mokyr 2002). Not only does novel scientific knowledge provide guidance in the process of invention

(Nelson 1982; Fleming and Sorenson 2004), it can also at times be directly instantiated as a new technology (Stokes 1997; Murray 2002). Yet, the creation of new scientific knowledge is not enough to ensure its development into novel inventions (Zucker, Darby, and Armstrong 2002; Murray and O'Mahony 2007). While both firms and universities produce scientific knowledge (Sauermann and Stephan 2010), a debate exists between those who believe that the academic environment is the most propitious to cumulative invention and those who believe the contrary.

2.1. The Entrepreneurial University view of technology spawning

The idea that the academic environment fosters cumulative invention is based on the observation that the diffusion of knowledge is central to the work and mission of university scientists. The ability to claim priority for an original contribution is a fundamental determinant of the allocation of honor in the scientific community and honor in turn shapes individual careers. In such a regime, appropriation does not occur through secrecy but through openness, sharing and claiming. Races and disputes are therefore common, since scientists have a strong incentive to rush and disseminate the knowledge that they have created (Merton 1957). This system pushes scientists to disclose quickly and to lower as much as they can the cost for others to access and reuse the new knowledge that they create. While firms might have incentives to stall the development of technologies that could cannibalize other sources of revenue, academic researchers can only gain from an increase in recognition. Similarly, the transparent evaluation of individual scientists and of their work through peer review lowers the uncertainties for R&D labs interested in absorbing knowledge or hiring personnel, thus decreasing the cost of learning and cumulative invention (Dasgupta and David 1994). The institutional arrangement ensuring the preservation, certification and distribution of scientific knowledge at a low cost for the cumulative innovator is sophisticated and includes both informal means such as conferences and peer networks, as well as formal mechanisms such as peer-reviewed publications and materials libraries such as biomedical resource centers (Furman and Stern 2011).

While firms can also be involved in the production of "open science" (Stephan 1996; Sauermann and Stephan 2010), industrial R&D is typically associated with a very different institutional environment. Because firms value the commercial logic of science based on exclusion and secrecy, scientists can earn higher wages if they accept to not be a part to the community of open science (Stern 2004). In commercial science, property over a new piece of scientific knowledge is not achieved through sharing, but through excluding. The incentive is not to make cumulative invention by others easier; it is to make it more costly. This sharp contrast is well illustrated by the case of Harvard's exclusive licensing of the OncoMouse patent to DuPont (Murray 2010). From its discovery in 1984 to its entry into the commercial world in 1988, geneticists were pushed by the generous material culture and strong normative pressure to freely exchange the mouse. This in turn enabled a fast expansion of cumulative innovation by other labs. However, after it received its exclusive license in 1988, DuPont introduced a number of restrictions on the exchange of the mouse, increasing its cost, prohibiting informal exchanges, introducing annual reporting requirements, and imposing reach-through on future discoveries made with the OncoMouse. These decisions caused an outrage in the scientific community and greatly decreased the rate of cumulative innovation based on the mouse's discovery.

In sum, the Entrepreneurial University view of technology spawning is based on the distinct institutional logics prevailing in industry and in academia. A number of researchers have argued in favor of this view (David and Dasgupta 1987; Zucker, Darby, and Brewer 1998; Henderson, Jaffe, and Trajtenberg 1998; Adams 2002; Owen-Smith and Powell 2004; Furman et al. 2005). In fact, it is well summarized in the following words: "since universities are in principle dedicated to the widespread dissemination of the results of their research, university spillovers are likely to be disproportionately large and may this be disproportionately important" (Henderson, Jaffe, and Trajtenberg 1998, 119).

2.2. The Ivory Tower view of technology spawning

In contrast, the Ivory Tower view of technology spawning emphasizes that the academic environment impedes cumulative innovation. The emphasis here is not on institutional norms, it is instead on the community of inventors and on the circulation of ideas in that community.

While academic institutions have the mission to disseminate their knowledge, individual scientists often lack the incentive—and sometimes the resources—to develop their ideas beyond the stage of proof of concept of early prototype (Jensen and Thursby 2001). While non-discoverers could potentially undertake cumulative innovation based on the new discovery, academic scientists are often reluctant to spend the time and resources necessary to transfer the tacit knowledge to outsiders (Zucker, Darby, and Armstrong 2002). They are also often uninformed about commercialization opportunities. In addition, the logistics of knowledge transfer involves contractual agreements which can be quite costly and complicated, and university scientists might not have access to the appropriate supporting infrastructure (Colyvas et al. 2002).

In contrast to scientists working in academia, firm scientists have strong incentives to develop new technologies based on the scientific knowledge that they create (Aghion, Dewatripont, and Stein 2008). Moreover, even if firms try to limit knowledge diffusion toward other firms by using for instance non-compete agreements, the effectiveness of these types of measures is limited (Marx, Strumsky, and Fleming 2009) and knowledge often flows between competing firms (Saxenian 1994). Structurally, universities often occupy peripheral positions in networks of innovative organizations (Powell et al. 2005) and scientific and inventor collaboration networks are similarly distinct (Murray 2002). Thus, scientists working at industrial R&D labs tend to primarily use external knowledge that they find from other firms rather than from academic institutions (Cohen, Nelson, and Walsh 2002).

The Ivory Tower view of technology spawning is based on the classic image that promising scientific knowledge might be sitting unused on the shelves of university labs. This view was well summarized by senator Birch Bayh, author of the famous Bayh-Dole Act, when he famously declared: “What sense does it make to spend billions of dollars each year on government-supported research and then prevent new developments from benefiting the American people because of dumb bureaucratic red tape?” (AUTM 2004, cited in Alridge & Audretsch 2010). Thirty years after the Bayh Dole act, university patenting has boomed (e.g. Mowery, Sampat, and Ziedonis 2002) and university knowledge is being increasingly translated into new technologies (Sampat, Mowery, and Ziedonis 2003; N. Hausman 2010). Yet recent studies have shown that many difficulties remain (Grimaldi et al. 2011).

3. EMPIRICAL APPROACH

3.1. The Challenge

The empirical challenge in testing the two hypotheses of technology spawning is straightforward. When observing the emergence of science-based invention, how can we gauge whether these stem from the intrinsic potential of the knowledge itself rather than from the characteristics of the environment of discovery? Universities are widely believed to conduct much more basic research than firms. As a consequence, the relevance of university research for invention tends to be indirect whereas firms produce directly relevant knowledge. A large number of studies have described the division of labor between firms and universities (Nelson 1986; Rosenberg and Nelson 1994; Rosenberg 1994; Klevorick et al. 1995; Mansfield 1995; Cohen, Nelson, and Walsh 2002).

The fundamental empirical challenge is therefore an identification problem. The risk is to conflate the marginal impact of the environment of discovery with the selection effect of knowledge into this environment. Knowledge with high potential for cumulative invention is more likely to emerge in firms rather than in universities. A simple comparison between different types of environments (e.g. university vs. industry) might therefore lead to biased results due to unobserved differences in technological potential. Ideally, the researcher would like to observe the potential of the new knowledge and to compare this potential to realized cumulative invention.

3.2. Paper-Twins

This paper proposes a novel empirical approach exploiting the existence of simultaneous discoveries operationalized as paper-twins. Paper twins are the dual instantiation of the same piece of new scientific knowledge in two distinct environments. The following example resulted from a discovery simultaneously made at UCSF in California and at SmithKline Beecham in England:

> **Caterina et al. (April 2000) "Impaired Nociception and Pain Sensation in Mice Lacking the Capsaicin Receptor."** ***Science***
> "The capsaicin (vanilloid) receptor VR1 is a cation channel expressed by primary sensory neurons of the "pain" pathway. Heterologously expressed VR1 can be activated by vanilloid compounds, protons, or heat (>43°C), but whether this channel contributes to chemical or thermal sensitivity in vivo is not known. Here, we demonstrate that sensory neurons from mice lacking VR1 are severely deficient in their responses to each of these noxious stimuli. VR1−/− mice showed normal responses to noxious mechanical stimuli but exhibited no vanilloid-evoked pain behavior, were impaired in the detection of painful heat, and showed little thermal hypersensitivity in the setting of inflammation. Thus, VR1 is essential for selective modalities of pain sensation and for tissue injury–induced thermal hyperalgesia."

> **Davis et al. (May 2000) "Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia."** ***Nature***
> "The vanilloid receptor-1 (VR1) is a ligand-gated, non-selective cation channel expressed predominantly by sensory neurons. VR1 responds to noxious stimuli including capsaicin, the pungent component of chilli peppers, heat and extracellular acidification, and it is able to integrate simultaneous exposure to these stimuli (…). Here we have disrupted the mouse

> VR1 gene using standard gene targeting techniques. (…) Although the VR1-null mice appeared normal in a wide range of behavioural tests, including responses to acute noxious thermal stimuli, their ability to develop carrageenan-induced thermal hyperalgesia was completely absent. We conclude that VR1 is required for inflammatory sensitization to noxious thermal stimuli but also that alternative mechanisms are sufficient for normal sensation of noxious heat."

These excerpts describe two sets of results obtained by examining the behavior of mice lacking a specific receptor (VR1). Both teams have found that mice in which the VR1 gene had been disrupted exhibit normal reactions to a variety of stimuli but become completely insensitive to one specific stimulus (carrageenan-induced thermal hyperalgesia). One of the team (Caterina et al) conducted its research within academia and the other team (Davis et al) in a firm. Both papers were submitted within a month (respectively January 18th 2000 and December 20th 1999). In short, the (nearly) simultaneous discovery of the capsaicin receptor in two different environments led the disclosure of the same new knowledge in two distinct papers.

We use simultaneous discoveries as an "experiment" from which it is possible to compare the relative impact of the academic and corporate environments on cumulative invention. Specifically, our empirical strategy exploits three key aspects of the phenomenon associated with the production of paper-twins:

a. since they disclose the same discovery, the knowledge disclosed in each of the paper-twins has intrinsically the same potential for follow-on inventions
b. since simultaneous discoveries emerge in different environments, the knowledge from each discovery might not actually be turned into new inventions at the same rate
c. citation and non-citation of each of the twin papers in the patent literature are a noisy but useful measure of the occurrence (or non-occurrence) of cumulative invention

3.2. Cumulative invention by the discoverers

Using simultaneous discoveries, it is possible to compare the extent to which discoverers engage in cumulative invention in academia and in firms. Empirically, our goal is not to estimate whether universities or firms get the "paired patent" (Murray 2002) on the newly discovered knowledge itself. Receiving the paired patent depends mostly on the exact timing of the discovery as well as on the patent application strategy. In contrast, the focus of this paper is on the long term use of the new knowledge as a springboard for invention. Since they discovered the same knowledge around the same time, the academic and industrial scientists in the dataset have the same opportunity to turn this knowledge into cumulative inventions. As a measure of discoverer invention, we use the number of patents in the patent literature that (1) originate from a discoverer and (2) that cite one of the discovery papers.

If the academic environment holds back cumulative invention by the discoverer, then the academic discovery team should receive fewer patents based on the discovery than the corporate co-discoverers. Since firms are known to conduct on average more applied research than universities, not accounting for a discovery's technological potential would bias the results. I can introduce paper-twins fixed-effects in order to examine the extent to which cumulative invention varies across discovery teams while keeping the discovery constant. Empirically, measuring cumulative invention using patent citations implies that we must account for its form as count data skewed to the right, calling for the use of a count model such as a fixed effect Poisson with quasi-maximum likelihood (i.e. "robust") estimates (J. Hausman, Hall, and Griliches 1984). In addition, USPTO patenting by the discoverer is likely to depend on whether the discovery was made in the US as well as on the number of discoverers. Our baseline empirical test for the impact of the academic environment on invention by the discovery team of paper *i* of twin *j* is therefore:

$$\#\, SELF\ PATENTS_{i,j} = f(\varepsilon_{i,j};\ \alpha_0 + \alpha_1 ACAD_i + \alpha_2 X_i + \gamma_j)$$

where γ_j is a paper-twin fixed effect, $\alpha_2 X_i$ is a vector of control variables including team size and location in the United States and $ACAD_i$ is a dummy variable equal to one if the principal investigator on the paper is based in academia and zero if he or she is based in industry. Because the dataset consists exclusively of simultaneous discoveries that involve at least one academic team and one industry team, we are able to identify differences in the rate of cumulative invention based on the same discovery in these two environments.

Discoverer invention is however an incomplete measure of a potential ivory tower effect. A number of researchers have noted that faculty patenting is only one channel through which cumulative invention might occur and that its relative importance is limited (Agrawal and Henderson 2002). It is possible that academic scientists systematically "outsource" cumulative invention to scientists from other organizations. Faculty patenting might therefore be a deceiving measure of cumulative invention since knowledge produced in universities might be turned into novel technologies by inventors who did not take part to the discovery. In order to get a deeper understanding of the potential ivory tower effect, it is therefore crucial to analyze the extent to which non-discoverer inventors are able to draw on scientific knowledge stemming from firms and universities.

3.3. Cumulative invention by non-discoverers

One crucial advantage of using paper-twins is that it allows the observation of cumulative inventions that do not cite one (or more) of the papers on which they build. As detailed in part 4, the incentive of the inventor is to cite the entire prior art of which he or she is aware. While citation to one of the twin papers indicates cumulative invention, failure to cite all twin papers can be interpreted as incomplete knowledge of the prior art by the cumulative inventor[2]. For each cumulative patent, we can contrast these instances of "failed" knowledge flow to the actual observed citations, or "realized" knowledge flow. We are thus able to estimate the impact of the environment of discovery—academia versus industry—on the rate of follow-on citations in the cumulative patent literature.

If the academic environment hinders cumulative invention by non-discoverers, the rate of citations to academic papers should be inferior to the rate of citations to their industry twin. Since unobserved characteristics of the inventor or invention (e.g. familiarity with the scientific literature) might be correlated with the origin of the scientific discovery, we use citing-patent fixed effects in order to avoid an omitted variable bias. The binary nature of the outcome variable could be modeled using a logistic regression with citing patent fixed effects. However, considering the small number of observations per citing patent, such a model would not be consistent. The well-known incidental parameter problem can be solved by using a conditional likelihood function instead of the usual maximum likelihood. We therefore carry out the estimation using a conditional logit model (Chamberlain 1982). As in the previous analysis, the realization of a citation might depend on characteristics of the environment of discovery such as whether the discovery was made in the US as well as on the number of discoverers. In this case, characteristics of the patent-paper dyad might also impact citation. These include for instance the number of years separating the discovery and the invention as well as the geographic distance separating the discoverers and the inventors. Our baseline empirical test for the impact of the

[2] Non-realized citations can be interpreted either as failed knowledge flow or as failed invention. Unfortunately, our dataset does not allow us to disentangle these two interpretations.

academic environment on the extent to which invention *k* has drawn knowledge from paper *i* of twin *j* is therefore:

$$CITATION_{kij} = f(\varepsilon_{kij};\ \alpha_0 + \alpha_1 ACAD_i + \alpha_2 X_{ki} + \gamma_{kj})$$

where γ_{kj} is a fixed effect for patent *k* citing discovery (paper twin) j, $\alpha_2 X_{ki}$ is a vector of control variables including characteristics of the papers as well as characteristics about the patent-paper dyad and $ACAD_i$ is a dummy variable equal to one if the corresponding author of the paper is based in academia and zero if he or she is based in industry. Because we observed two sources—academia and industry—for the same knowledge, we are able to measure the propensity to learn from each environment independently of the technological potential of the knowledge produced.

3.4. Nuances of the effect of the environment

In order to gain a finer understanding of the mechanism at play in our data, we also explore a number of more nuanced hypotheses about cumulative invention by both the discoverers and non-discoverers. First, discoverer-inventors from academia are less likely to have the appropriate industrial and market-knowledge about the type of inventions that could be successful (Gittelman and Kogut 2003). We therefore expect discoverer-inventors from academia to produce lower impact patents than discoverer-inventors from industry. Second, we expect that connection to a network of inventors is a driver of cumulative invention. Thus, the size of the team of industry scientists and the geographic proximity between the discovery and the inventor ought to increase the relative impact of the industry environment on cumulative invention. We also expect the difference between the rate of cumulative invention based on academic and corporate discoveries to decrease overtime as information spreads about the simultaneous discovery.

4. THE DATA

4.1. Sample definition

The data for this study is based on the first automatically and systematically collected dataset of simultaneous discoveries. The full dataset consists in 1,246 papers disclosing 578 discoveries and operationalized as 720 paper twins[3] published between 1970 and 2009. The algorithm used to build this dataset is based on the insight that two papers disclosing the same simultaneous discovery are systematically cited together in the follow-on literature, not only in the same papers, but also in the same parenthesis, or adjacently. The algorithm is summarized in Figure 1 and described with more details in a separate paper (Bikard 2012).

Insert Figure 1 about here

Our empirical work relies on the fact that paper-twins are indeed simultaneous discoveries and have therefore inherently the same potential for cumulative invention. Observed variance in the citation rate of two twin papers in the patent literature ought therefore to be due to the different environments in which the research took place rather than on differences in the discovery itself. We check this comparability assumption in several ways[4]. First, we examine the number of months separating the publications of two twin papers. As noted above, the algorithm did not match articles on simultaneity beyond ensuring that the calendar years of publications were no more than one year apart. If two alleged paper twins were not really the same, one would expect them to be on average six months apart or more. The 720 paper twins were in fact published on average 1.8 months apart, a lag considerably shorter than the average time between paper submission and publication. In fact, 373 pairs of twins were published the exact same month and 267 of them were published in the same issue of the same journal. We then verify the semantic similarity of two twin papers by using the Pubmed related citation algorithm. If the twins were not very closely related, they should not be using the same words and should therefore be ranked far from each other. Pubmed ranks two papers of the same twin right next to each other 42% of the time. The rank different is inferior to 10 for 90% of the twins[5]. Finally, the best test is probably to collect the opinion of the discoverers themselves. We selected randomly 10 discoverers and asked them to describe the discovery process. 9 of them told us that it was an instance of simultaneous discovery without us asking. After we asked the tenth person why he did not mention the twin paper, he asserted angrily that he deserved all the credit and that his idea had been stolen. Of course, the fact that two teams have published twin papers at the same

[3] In the data a triplet appears as 3 paper twins, a quadruplet as 6 paper twins

[4] This analysis was done for the entire dataset consisting of 720 paper-twins (578 simultaneous discoveries)

[5] Rank difference calculated after dropping articles that are published more than a calendar year apart

time does not mean that they conducted the exact same experiment or that they interpreted their results in the exact same way. It also does not mean that both discoveries were independent of one another. In fact, interviews with discoverers uncovered a number of cases in which one team accused the other of idea theft. However, the fact that the community of experts ruled that credit ought to be shared between two teams does indicate (1) that both teams are widely believed to have had the capability to make the discovery and (2) that each team has provided convincing evidence supporting their claim to priority.

The core of the analysis performed here is based on a subset consisting in 90 scientific publications disclosing 39 simultaneous discoveries having involved at least one industry-based team and one team based in a public research organization. We tracked cumulative invention by examining the citations of each of the 90 papers in the patent literature. We used a web crawler that searched the title of each paper and output information about each of the patents that included the article's title in its body. References in patents are important since they define the scope of the claimed novelty. As such, they are the responsibility of the inventor, the attorney and the examiner. In the US, the applicant has a strong incentive to disclose all prior art that he or she is aware of because failure of doing so can lead to patent invalidation, a rule known as the doctrine of "Inequitable Conduct." Citations in the patent literature are an imperfect measure of knowledge diffusion because not all innovations are patented (Cohen, Nelson, and Walsh 2000), not all knowledge flows are cited or citable (Griliches 1990), citations are at time used strategically (Lampe 2010) and a number of them are added by the examiner (Alcácer and Gittelman 2006). Yet, they are a readily available, comprehensive and well understood measure of knowledge dissemination and are therefore widely used (Jaffe, Trajtenberg, and Henderson 1993; Narin, Hamilton, and Olivastro 1997; Henderson, Jaffe, and Trajtenberg 1998; Gittelman and Kogut 2003; Sorenson and Fleming 2004). In addition, our particular setting presents two characteristics that ought to attenuate some of the concerns associated with this measure. First, we are studying life sciences, an area in which patents are widely used and strategic citations is limited (Lampe 2010). Second, we are studying citations to scientific papers, which tend to be less added by the examiner, less strategically used, and overall a better measure of knowledge diffusions than patent citations to other patents(Cohen, Nelson, and Walsh 2002; Roach and Cohen 2011).

As apparent in Table 1, the data is drawn from several sources. Data about each publication comes from ISI Web of Science and Scopus. Details about the corresponding author come from an analysis of the publications themselves. Patent citation data (through May 2011) and information about each citing patent were collected using Google Patents.

Insert Table 1 about here

4.2. Summary Statistics

Table 2 and Figure 2 use the entire dataset of 1,246 papers and 578 simultaneous discoveries and present evidence that university science tends to be more basic—at least as measured by the patent citations that simultaneous discoveries receive. The two graphs on top of Figure 2 show the different rates of yearly patent citations for non-matched academic and industry publications. Clearly, the average academic paper in the dataset receives far fewer patent citations than the average industry publication. The two bottom graphs show the same results for the "matched sample" i.e. the discoveries made simultaneously by an academic and an industrial team. As apparent from the graph, the difference in citation rate is much smaller. Table 2 similarly compares the unmatched and the matched samples and presents the same result. Thus, absent a close control for the fundamental aspect of academic research, comparisons of academic and industrial publications would vastly bias the results against the Entrepreneurial University view of technology spawning. While the analysis presented here focuses exclusively on the rate of cumulative invention, it is striking to also note in Table 2 that patents referring to firm papers are largely more cited than those referring to the academic twin, therefore providing additional evidence in support of the Ivory Tower view.

Insert Table 2 & Figure 2 about here

For the main analysis, we restrict our dataset to the "matched sample" only. Table 3 reports summary statistics. The sample consists of 39 simultaneous discoveries disclosed in 90 scientific publications and that have received 533 citations in the patent literature. In order to investigate discoverer invention, we measure # SELF PATENTS, the number of cumulative inventions made by the discoverers themselves. There are 91 such patents in the dataset that cite

18 papers—or 13 simultaneous discoveries. In contrast, the examination of non-discoverer invention concentrates on each potentially citing patent. The CITATION variable takes a value of 1 if the citation has taken place and 0 otherwise. The dataset includes 742 potential citations of which 442 are realized. One can also note that the 39 simultaneous discoveries were made between 1994 and 2008 and that 58 out of the 90 papers had their corresponding addresses in the US. The average number of authors per paper is 13.9 but one paper-twin about the sequencing of chromosomes of the plant Arabidopsis thaliana (both published in the same issue of Nature in 1999) included a 216-authored and a 37-authored paper.

Insert Table 3 about here

ACADEMIA is a dummy variable equal to one for all papers whose corresponding author was based in a university or public research organization. In our dataset, the 39 simultaneous discoveries took place in 41 public research institutions and 25 firms. The most common public research institutions in the data are Harvard University (4 papers), UT Houston (2 papers) and Stanford University (2 papers) and the most common firms are Genentech (6 papers), GSK (5 papers) and Amgen (4 papers). The 533 citations that the papers have received in the patent literature represent 313 unique patents, 75% of them were assigned to firms and 25% of them were assigned to academic institutions. While two twin-papers are consistently cited together in the scientific literature, the same is not true in the patent literature. The intersection of the forward citations of paper twins in the patent literature is only 21% of the union.

The relationship between ACADEMIA and the two main dependent variables # SELF PATENTS and CITATION are plotted in Figure 3 and 4 respectively. Figure 3 shows that industry papers are clearly associated with a higher number of discoverer patent citations (z=1.64 and p=0.10 in two-sample Wilcoxon Mann-Whitney test). Similarly, the fact that the share of realized citations is higher from industry papers is apparent in Figure 4 (z=1.87 and p=0.19 in McNemar chi-square test for matched pairs).

Insert Figure 3 and 4 about here

----

5. EMPIRICAL ANALYSIS

The empirical analysis involves two stages. Step 1 focuses on the count of patents by one of the discoverers and employs a fixed effect Poisson estimator with conditional maximum likelihood. Step 2 estimates the propensity that a patent will cite a discovery paper and uses a conditional logistic model. Robust standard errors are respectively clustered at the level of the twin and at the level of the citing patent.

5.1. Cumulative invention by the discoverers

Table 4 presents the results of the discoverer invention analysis. It includes simultaneous discovery fixed-effect so as to control for unobserved characteristics of each discovery. The first two columns use as dependent variables the count of discoverer patents and the two right-hand-side columns present the same regressions but use as dependent variable the count of cumulative patents assigned to the organization of discovery (rather than the individual discoverer). The results are in line with the Ivory Tower view of technology spawning. They show that cumulative invention by the discoverers is much weaker in academia as compared to firms. In fact, firm discoverers produce on average over three times more patents than academic discoverers based on the same discovery. Interestingly, the coefficient is even stronger for the organization of discovery than it is for the individual discoverers. This result indicates that cumulative invention in firms does not necessarily involve the discoverer. The same is apparently less true in universities.

Insert Table 4 about here

5.2. Cumulative invention by non-discoverers

We turn to an analysis of the propensity of non-discoverer inventors to draw upon knowledge produced in academia or in industry. Paper-twins can be used to observe the non-dissemination of knowledge since all patents citing at least one of the twins could potentially have cited all of them. Table 5 presents the results of the baseline conditional logistic regressions, estimating the realization of citations as a function of whether the discovery paper stems from the academia. The negative impact of the academic environment on cumulative invention by non-discoverers appears modest (10-20%) but statistically significant and robust to the inclusion of a number of control variables, including characteristics of the paper such as number of authors, US-based, and characteristics of the patent-paper dyad such as time lag and geographic distance. Taken together, these results bring further support to the Ivory Tower view of technology spawning. Cumulative inventors are more likely to draw on a piece of scientific knowledge if it emerges in a firm as compared to a university.

Insert Table 5 about here

5.3. Nuances of the effect of the environment

The remaining two tables provide a more granular view of the negative impact of the academic environment on cumulative invention by non-discoverers. Table 6 presents an analysis of the variation in the effect as a function of the number of authors, the time after discovery, whether the inventor and the discoverer were based in the same country and whether they were located geographically close to each other. In line with the explanation that the difference in cumulative invention is driven by the denser connection of firm scientists in the inventor community, we find that the negative impact of the academic environment increases with the size of the discovery team. This same negative effect also seems to be particularly salient in the years immediately following the discovery and to become weaker overtime. Similarly, the negative effect appears weak in the instance in which the discoveries and the inventions were made in different countries and is stronger when both happened in the same country. Finally, the last column of the table shows that the negative effect of the academic environment decreases the further one is from the place of discovery. Predicted values from this regression are plotted in Figure 5. This statistically significant interaction effect is particularly telling since it indicates that the observed effect cannot be explained by the potential existence of a norm leading to the citation of firm papers rather than academic ones in patents.

Insert Table 6 and Figure 5 about here

Finally, table 7 splits the sample between cumulative inventors from academia and those from firms. The effect of the environment of discovery might differ for both types of inventors since academic inventors might be more familiar with the work of other university scientists than firms' researchers. On the contrary, table 6 indicates that the negative impact of the academic environment seems just as strong for inventors from universities and firms. We do not find evidence that knowledge circulates better within the "ivory tower".

Insert Table 7 about here

One could be worried that our results might be driven not by the flow knowledge but instead by some norm or some strategic decision to cite corporate rather than academic discoveries when both are available. While we cannot entirely disprove this possibility, we can note that such a theory would not be consistent with our finding that the negative impact of the academic environment increases with geographic proximity (Figure 5). Similarly, the finding that the effect remains the same for academic inventors (Table 7) suggests that this citation patterns is not purely driven by the concern of getting sued. Finally, we conducted 5 interviews with inventors in order to inquire about their citation decision. All of them affirmed that they cited all the papers that they were aware of and that they regarded as most relevant. Inventors typically justified the non-citation of the twin by mentioning "lack of awareness," "lack of time," and in one case "lack of clarity" of the twin paper.

6. DISCUSSION

This paper uses a novel empirical strategy to test the relative impact of the academic and corporate environments on follow-on cumulative invention, a process that we called technology spawning. The Entrepreneurial University view of technology spawning observes that universities aim at disseminating the knowledge that they create and concludes that the academic environment increases the rate of cumulative invention based on a scientific discovery. The Ivory Tower view of technology spawning, on the other hand, points to the lack of connection between universities and the community of inventors. It argues that the academic environment has a negative impact on cumulative invention.

The empirical approach presented here exploits the existence of simultaneous discoveries and their instantiation as paper-twins. Because simultaneous discoveries can emerge on both sides of the academia-industry boundary, it is possible to examine the same piece of new knowledge in two different institutional settings. This paper uses the first systematically and automatically generated dataset of simultaneous discoveries, including 578 instances. The core of the study focuses on 39 such discoveries that involved at least one team from academia and one team from industry. Analysis of the citations of the twin papers in the patent literature indicates that the academic environment has a negative impact on cumulative invention.

We propose that science-based cumulative invention—or technology spawning—results from two distinct processes. First, cumulative invention can be undertaken by one of the discoverers. Second, non-discoverers might also use the new knowledge from the discovery in their own invention effort. We find that the academic environment has a significant negative impact on the rate of cumulative invention in both cases. However, the magnitude of this "ivory tower effect" is much stronger in the case of the discoverer's inventions. The rate of cumulative invention by the discoverer is over three times higher in industry than in academia. In addition, non-discoverers inventors are 10-20% more likely to draw their knowledge from firms rather than from academic institutions.

These findings should be interpreted carefully. First, we have not identified the underlying mechanism by which scientific knowledge in academia tends less to be turned into new inventions. We cannot disentangle whether the low rate of technology spawning from university science comes about through a lack of awareness of a specific discovery's potential, lower rates of diffusion in the inventor community, or even through some underlying unobserved cost of drawing knowledge from an academic lab rather than from a firm. Second, since this study is based on 39 simultaneous discoveries made by 90 teams, one could potentially question the generalizability of our results. Simultaneous discoveries that involve both firms and universities might be a special case of discoveries since they involve clearly more authors and receive more patent citations than the simultaneous discoveries that do not involve firms (see Table 2). In addition, one could argue that in the case of a simultaneous discovery, the attention of the cumulative inventors simply shifts toward the industry twin. On the other hand, considering that the large majority of the discoveries in the dataset are from the life sciences—a field in which university-industry collaboration is particularly intense—the dataset might constitute a lower boundary of the propensity of the academic environment to trap commercially relevant knowledge within its walls.

The evidence that knowledge might to some extent be trapped inside universities supports the widespread idea that academic research might be somehow "under-exploited"—at least as compared to discoveries made in firms. This could also explain why firms that collaborate with universities tend to be more innovative (e.g. Cockburn and Henderson 1998). However, the implications of these findings for firms and policy-makers would depend considerably on the specific mechanism underlying these results. The observed differences in cumulative invention might originate from the cost to access technologically relevant information or from the choice of inventors to learn from firms rather than from universities. Ongoing qualitative work by the author should provide a more precise identification of the mechanism at play as well as deeper insights into the practical implications of this research.

One should also note that the evidence presented here, that research remains "trapped" inside the ivory tower, captures only one aspect of the impact of the academic environment on knowledge dissemination. This paper starts when the discovery process stops, and therefore does not explore the antecedents of knowledge creation, including the ability to build on other scientists' shoulders (Furman and Stern 2011). Without a detailed accounting of the size of other (positive) effects of the academic environment on knowledge dissemination, it is impossible to calculate the optimal innovation policy approach towards university science.

The academic research environment can be portrayed as an ivory tower. On the one hand, research conducted there tends to be more basic and less directly relevant to science-based invention. On the other hand, even the relevant research done there is less likely to be turned into inventions than it would have, had it been conducted in the private sector. By focusing on simultaneous discoveries, it is possible for the first time to observe the non-occurrence of otherwise possible cumulative inventions. The list of potential drivers and obstacles to technology spawning is long. Considering the growing desire to see publicly funded scientific research contribute to the economy through its translation into novel technologies, the use of simultaneous discovery as "knowledge twins" presents tremendous opportunities for future research.

REFERENCES

Adams, James D. 2002. "Comparative Localization of Academic and Industrial Spillovers." *Journal of Economic Geography* 2 (3): 253–278.

Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein. 2008. "Academic Freedom, Private-Sector Focus, and the Process of Innovation." *The RAND Journal of Economics* 39 (3) (October 1): 617–635.

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

Alcácer, Juan, and Michelle Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88 (4): 774–779. doi:i: 10.1162/rest.88.4.774</p>.

Bikard, Michaël. 2012. "A Systematic and Automated Method to Build Datasets of Simultaneous Discoveries." *MIT Sloan Working Paper*.

Chamberlain, G. 1982. *Analysis of Covariance with Qualitative Data*. National Bureau of Economic Research Cambridge, Mass., USA.

Cockburn, Iain M., and Rebecca M. Henderson. 1998. "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery." *The Journal of Industrial Economics* 46 (2) (June): 157–182.

Cohen, Wesley M., and Daniel A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35 (1) (March): 128–152. doi:10.2307/2393553.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. *Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not)*. National Bureau of Economic Research.

———. 2002. "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48 (1): 1–23.

Colyvas, Jeannette, Michael Crow, Annetine Gelijns, Roberto Mazzoleni, Richard R. Nelson, Nathan Rosenberg, and Bhaven N. Sampat. 2002. "How Do University Inventions Get into Practice?" *Management Science* 48 (1) (January): 61–72.

Dasgupta, Partha, and Paul A. David. 1994. "Toward a New Economics of Science." *Research Policy* 23 (5) (September): 487–521. doi:10.1016/0048-7333(94)01002-1.

David, PA, and P. Dasgupta. 1987. "Information Disclosure and the Economics of Science and Technology'." *Arrow and the Ascent of Modern Economic Theory, Macmillan, London*.

Fleming, Lee, and Olav Sorenson. 2004. "Science as a Map in Technological Search." *Strategic Management J* 25: 909–928.

Furman, Jeff, M.K. Kyle, I. Cockburn, and R.M. Henderson. 2005. "Public & Private Spillovers, Location and the Productivity of Pharmaceutical Research." *Annales d'Économie Et De Statistique*: 165–188.

Furman, Jeff, and Scott Stern. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." *American Economic Review* 101 (5): 1933–1963.

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) (April 1): 366–382.

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* 28 (4) (December): 1661–1707. doi:10.2307/2727442.

Grimaldi, Rosa, Martin Kenney, Donald S. Siegel, and Mike Wright. 2011. "30 Years After Bayh–Dole: Reassessing Academic Entrepreneurship." *Research Policy* 40 (8) (October): 1045–1057. doi:10.1016/j.respol.2011.04.005.

Hausman, J., B.H. Hall, and Z. Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R & D Relationship." *Econometrica: Journal of the Econometric Society*: 909–938.

Hausman, N. 2010. "Effects of University Innovation on Local Economic Growth and Entrepreneurship."

Henderson, Rebecca M., Adam B. Jaffe, and Manuel Trajtenberg. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." *Review of Economics and Statistics* 80 (1) (February): 119–127. doi:10.1162/003465398557221.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca M. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3) (August): 577–598. doi:10.2307/2118401.

Jensen, Richard, and Marie Thursby. 2001. "Proofs and Prototypes for Sale: The Licensing of University Inventions." *The American Economic Review* 91 (1) (March): 240–259.

Klevorick, Alvin K., Richard C. Levin, Richard R. Nelson, and Sidney G. Winter. 1995. "On the Sources and Significance of Interindustry Differences in Technological Opportunities." *Research Policy* 24 (2) (March): 185–205. doi:10.1016/0048-7333(93)00762-I.

Lampe, Ryan. 2010. "Strategic Citation." *Review of Economics and Statistics*. doi:10.1162/REST_a_00159. http://dx.doi.org/10.1162/REST_a_00159.

Mansfield, Edwin. 1995. "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing." *The Review of Economics and Statistics* 77 (1) (February): 55–65.

Marx, M., D. Strumsky, and L. Fleming. 2009. "Noncompetes and Inventor Mobility: Specialists, Stars, and the Michigan Experiment." *Management Science* 55: 875–889.

Merton, Robert K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22 (6) (December): 635–659.

———. 1961. "Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science." *Proceedings of the American Philosophical Society* 105 (5) (October 13): 470–486.

Mokyr, J. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Mowery, David C., Bhaven N. Sampat, and Arvids A. Ziedonis. 2002. "Learning to Patent: Institutional Experience, Learning, and the Characteristics of U. S. University Patents After the Bayh-Dole Act, 1981-1992." *Management Science* 48 (1) (January 1): 73–89.

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): 341–388. doi:10.1086/653599.

Murray, Fiona, and S. O'Mahony. 2007. "Exploring the Foundations of Cumulative Innovation: Implications for Organization Science." *Organization Science* 18 (6): 1006–1021.

Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro. 1997. “The Increasing Linkage Between U.S. Technology and Public Science.” *Research Policy* 26 (3) (October): 317–330. doi:16/S0048-7333(97)00013-9.

Nelson, Richard R. 1982. “The Role of Knowledge in R&D Efficiency.” *The Quarterly Journal of Economics* 97 (3). The Quarterly Journal of Economics: 453–70.

Nelson, Richard R. 1986. “Institutions Supporting Technical Advance in Industry.” *The American Economic Review* 76 (2) (May 1): 186–189.

Owen-Smith, J., and W. W Powell. 2004. “Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community.” *Organization Science* 15 (1): 5–21.

Powell, Walter W., Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith. 2005. “Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences.” *American Journal of Sociology* 110 (4) (January 1): 1132–1205. doi:10.1086/421508.

Roach, Michael, and Wesley M. Cohen. 2011. “Lens or Prism? A Comparative Assessment of Patent Citations as a Measure of Knowledge Flows from Public Research.” *SSRN eLibrary* (July 15). http://papers.ssrn.com.libproxy.mit.edu/sol3/papers.cfm?abstract_id=1911965.

Rosenberg, Nathan. 1994. *Exploring the Black Box: Technology, Economics, and History*. Cambridge Univ Pr.

Rosenberg, Nathan, and Richard R. Nelson. 1994. “American Universities and Technical Advance in Industry.” *Research Policy* 23 (3) (May): 323–348. doi:10.1016/0048-7333(94)90042-6.

Sampat, Bhaven N., David C. Mowery, and Arvids A. Ziedonis. 2003. “Changes in University Patent Quality After the Bayh-Dole Act: a Re-examination.” *International Journal of Industrial Organization* 21 (9) (November): 1371–1390. doi:10.1016/S0167-7187(03)00087-0.

Sauermann, H., and P. E Stephan. 2010. *Twins or Strangers? Differences and Similarities Between Industrial and Academic Science*. National Bureau of Economic Research.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press.

Sorenson, Olav, and Lee Fleming. 2004. “Science and the Diffusion of Knowledge.” *Research Policy* 33 (10) (December): 1615–1634. doi:10.1016/j.respol.2004.09.008.

Stephan, Paula E. 1996. “The Economics of Science.” *Journal of Economic Literature* 34 (3) (September): 1199–1235.

Stern, Scott. 2004. “Do Scientists Pay to Be Scientists?” *Management Science* 50 (6) (June 1): 835–853. doi:10.1287/mnsc.1040.0241.

Stokes, D. E. 1997. *Pasteur’s Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press.

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) (January 1): 138–153.

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) (March): 290–306.

TABLE 1. Variables and definitions

Variable	Definition	Source
Publication characteristics		
PAPER-TWIN$_k$	Dummy variable for each pair of paper twins	Matching algorithm
ACADEMIA$_i$	Dummy variable equal to 1 if the corresponding author of article i is in a university or a government organization; 0 otherwise	Paper itself
US AUTHOR$_i$	Dummy variable equal to 1 if the corresponding author of article i is in the US; 0 otherwise	Paper itself
PUBLICATION YEAR$_i$	Year in which article i is published	WoS
# AUTHORS$_i$	Count of the number of authors of article i	WoS
Patent characteristics		
CITED BY PATENT	Dummy variable equal to 1 if article i is cited by a patent issued by the USPTO prior to May 2011; 0 otherwise	Google Patent
TOTAL PATENTS$_i$	# of patents citing article i issued before May 2011	Google Patent
UNIVERSITY ASSIGNEE	Percentage of assignees that are universities or a government organizations	USPTO
US INVENTOR	Dummy variable equal to 1 if the corresponding author of article i is in the US; 0 otherwise	USPTO
APPLICATION YEAR	Year of patent application to USPTO	USPTO
# CITATIONS RECEIVED	Number of citations received by patent j in the patent literature by May 2011	USPTO
Citation characteristics		
CITATION$_{ij}$	Dummy variable equal to 1 if article i is cited in patent j; 0 otherwise	Google Patent
CITATION LAG$_{ij}$	APPLICATION YEAR$_j$ - PUBLICATION YEAR$_i$	USPTO; ISI Web of Science (WoS)
GEOG DISTANCE$_{ij}$	Distance, in miles, between the cities of the address of publication i's corresponding author and patent j's first inventor (Law-of-Cosines-based calculation)	Harvard IQSS patent database; itouchmap.com
SAME COUNTRY$_{ij}$	Dummy variable equal to 1 if the corresponding author of article i is located in the same country as the first inventor of patent j; 0 otherwise	USPTO; WoS
SELF PATENT$_{ij}$	Dummy variable equal to 1 if patent j was assigned to an organization present in publication i's address field	USPTO; WoS

TABLE 2. Means conditional on discovery environment

	ALL PAPERS		TWIN ACROSS UNIVERSITY-INDUSTRY BOUNDARY	
	Univ. Paper	Firm Paper	Univ. Tower	Firm Paper
# Publications	1195	51	49	41
Patent citation characteristics				
# SELF PATENTS	0.12	1.92	0.43	1.71
# TOTAL PATENTS	1.61	9.96	4.80	6.90
AV. # OF CITES TO PATENTS	1.25	2.51	1.29	2.22
Publication characteristics				
US AUTHOR	0.61	0.64	0.61	0.65
# AUTHORS	7.25	12.73	14.47	13.27
PUBLICATION YEAR	2001.00	1999.71	2000.61	2000.24

TABLE 3. Means and standard deviations

Variable	N	Mean	Std. Dev.	Min	Max
Twin characteristics (N=39)					
# PAPERS PER TWIN	39	2.31	0.52	2	4
% TWINS CITED IN SELF PATENTS	39	0.33	0.48	0	1
# SELF CITING PATENTS PER TWIN	39	2.33	6.42	0	38
% TWINS CITED IN 3rd PARTY PATENTS	39	0.54	0.51	0	1
# 3rd PARTY PATENTS PER TWIN	39	9.64	15.80	0	57
Publication characteristics (N=90)					
ACADEMIA$_i$	90	0.54	0.50	0	1
US AUTHOR$_i$	90	0.64	0.48	0	1
# AUTHORS$_i$	90	13.92	24.06	2	216
PUBLICATION YEAR$_i$	90	2000.44	3.65	1994	2008
CITED BY SELF PATENT	90	0.20	0.40	0	1
# SELF PATENTS CITATIONS	90	1.01	3.26	0	26
CITED BY 3rd PARTY PATENT	90	0.44	0.50	0	1
# 3rd PARTY PATENTS CITATIONS	90	4.18	7.33	0	34
Discoverers' patent characteristics (N=91)					
% ACADEMIC ASSIGNEE	91	0.21	0.40	0	1
% US INVENTOR	91	0.92	0.27	0	1
APPLICATION YEAR	91	2003.59	3.20	1996	2009
# CITATIONS RECEIVED	91	3.75	5.54	0	37
3rd Party Patent-Paper dyad characteristics: Actual Citations (N=442)					
CITATION TO UNIV PAPER	442	0.49	0.50	0	1
TIME LAG (YEARS)	442	4.25	3.05	-2	15
SAME COUNTRY	442	0.69	0.46	0	1
GEOGRAPHIC DISTANCE (MILES)	442	2127.72	1857.51	0	9297.4
3rd Party Patent-Paper dyad characteristics: Non-Citations (N=300)					
CITATION TO UNIV PAPER	300	0.65	0.48	0	1
TIME LAG (YEARS)	300	4.11	3.10	-2	14
SAME COUNTRY	300	0.54	0.50	0	1
GEOGRAPHIC DISTANCE (MILES)	300	2543.74	1954.87	0	9728.2

TABLE 4. Impact of the discovery environment on discoverer invention

	FIXED EFFECT POISSON QML (level: simultaneous discovery)			
	DV = # PAT BY DISCOVERER		DV = # PAT BY DISCOVERY ORG	
	Marginal impact; no control	Marginal impact w/ controls	Marginal impact; no control	Marginal impact w/ controls
ACADEMIA	0.267***	0.268***	0.212***	0.258***
	(0.08)	(0.07)	(0.06)	(0.08)
Discovery team characteristics				
US AUTHOR		4.453		4.956e+07***
		(7.89)		(4.58e+07)
# AUTHORS		1.013		1.182
		(0.01)		(0.12)
# of observations	32	32	32	32
Paper-twin FE	13	13	13	13

Values are incident rate rations; robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 5. The impact of the discovery environment on cumulative invention by non-discoverers

	CONDITIONAL LOGIT (level: citing patent)		
	Dependent Variable = CITATION (dummy); Self-cites excluded		
	Baseline marginal impact; no control	Marginal impact; w/ discovery controls	Marginal impact; w/ discovery and dyad controls
ACADEMIA	0.580***	0.697**	0.576***
	(0.09)	(0.11)	(0.11)
Discovery team characteristics			
US AUTHOR		1.666**	0.665
		(0.43)	(0.27)
# AUTHORS		0.995*	0.998
		(0.00)	(0.00)
Patent-Paper Dyad characteristics			
CITATION LAG			1.136
			(0.39)
SAME COUNTRY			5.328***
			(2.54)
LOG (GEOG DISTANCE)			0.996
			(0.07)
# of observations	483	483	483
Patent-twin dyads FE	206	206	206

Values are odd ratios; Robust standard error in parenthesis are adjusted for 206 clusters (patent-twin dyads)

*** p<0.01, ** p<0.05, * p<0.1

TABLE 6. Variations in environment effect on cumulative invention by non-discoverers

	CONDITIONAL LOGIT (level: citing patent) Dependent Variable = CITATION (dummy); Self-cites excluded			
	Team size and academia effect	Citation lag and academia effect	Same country and academia effect	Geographic distance and academia effect
ACADEMIA	0.862	0.467***	0.722	0.103**
	(0.39)	(0.14)	(0.21)	(0.10)
Discovery team characteristics				
US AUTHOR	0.737	0.674	0.722	0.696
	(0.31)	(0.27)	(0.29)	(0.28)
# AUTHORS	1.025	0.998	0.997	0.997
	(0.03)	(0.00)	(0.00)	(0.00)
Patent-Paper Dyad characteristics				
CITATION LAG	1.125	1.072	1.167	1.061
	(0.38)	(0.37)	(0.40)	(0.37)
SAME COUNTRY	5.008***	5.295***	5.989***	4.570***
	(2.41)	(2.47)	(2.87)	(2.21)
LOG (GEOG DISTANCE)	0.995	0.997	0.994	0.823
	(0.07)	(0.07)	(0.07)	(0.12)
Interactions				
ACADEMIA*# AUTHORS	0.976			
	(0.02)			
ACADEMIA*CITATION LAG		1.049		
		(0.05)		
ACADEMIA*SAME COUNTRY			0.665	
			(0.27)	
ACADEMIA*LOG (GEOG DISTANCE)				1.270*
				(0.17)
# of observations	483	483	483	483

Values are odd ratios; Robust standard error in parenthesis are adjusted for 1244 clusters (patent-twin dyads)

*** p<0.01, ** p<0.05, * p<0.1

TABLE 7. Environment effect on academic vs. corporate non-discoverer inventors

	CONDITIONAL LOGIT (level: citing patent) Dependent Variable = CITATION (dummy); Self-cites excluded			
	Academia Effect: academic patents only		Academia Effect: corporate patents only	
ACADEMIA	0.556*	0.587	0.664**	0.526***
	(0.18)	(0.21)	(0.13)	(0.13)
Discovery team characteristics				
US AUTHOR	2.152	0.707	1.860*	0.993
	(1.03)	(0.62)	(0.62)	(0.53)
# AUTHORS	1.002	0.998	0.992**	0.998
	(0.01)	(0.01)	(0.00)	(0.00)
Patent-Paper Dyad characteristics				
LOG (DISTANCE)		0.113**		1.704
		(0.12)		(0.72)
SAME COUNTRY		2.571		3.935**
		(2.75)		(2.52)
CITATION LAG		0.877		0.948
		(0.20)		(0.08)
# of observations	117	117	332	332
Patent-twin dyads FE	54	54	137	137

Values are odd ratios; Robust standard error in parenthesis are adjusted for patent-level clusters

*** p<0.01, ** p<0.05, * p<0.1

FIGURE 1.An automated and systematic method to generate a list of simultaneous discoveries

Step 1: Collection of ISI Web of Knowledge data on all research articles from the 15 non-review scientific publications having the highest Journal Impact Factor
(42,106 publications)

↓

Step 2: Using Pubmed and CrossRef, verify the type of article and the complete author list of each of the 1,294,357 references online.
(744,583 unique references)

↓

Step 3: Generation of a database of pairs of all references (a) co-cited at least once, (b) written no more than 1 year apart, (c) having no overlapping author, (d) in which at least 5 citations for each reference are observed in the dataset of citing articles.
(17,050,914 pairs considered; 449,417 pairs selected)

↓

Step 4: Computation of the Jaccard co-citation coefficient for all pairs of references (intersection over the union of forward citations). Highly skewed distribution with a long tale of pairs that are consistently cited in the same papers.

↓

Step 5: Selection of the 2,320 pairs with co-citation coefficient superior to 50% and run a parsing algorithm on all the co-citing articles. Out of these pairs the parsing algorithm could analyze 3 co-citing publications or more in 1,825 cases; 720 pairs have been cited adjacently in 100% of the co-citing articles

FIGURE 2. Citation rates in patents: academic vs. industry papers

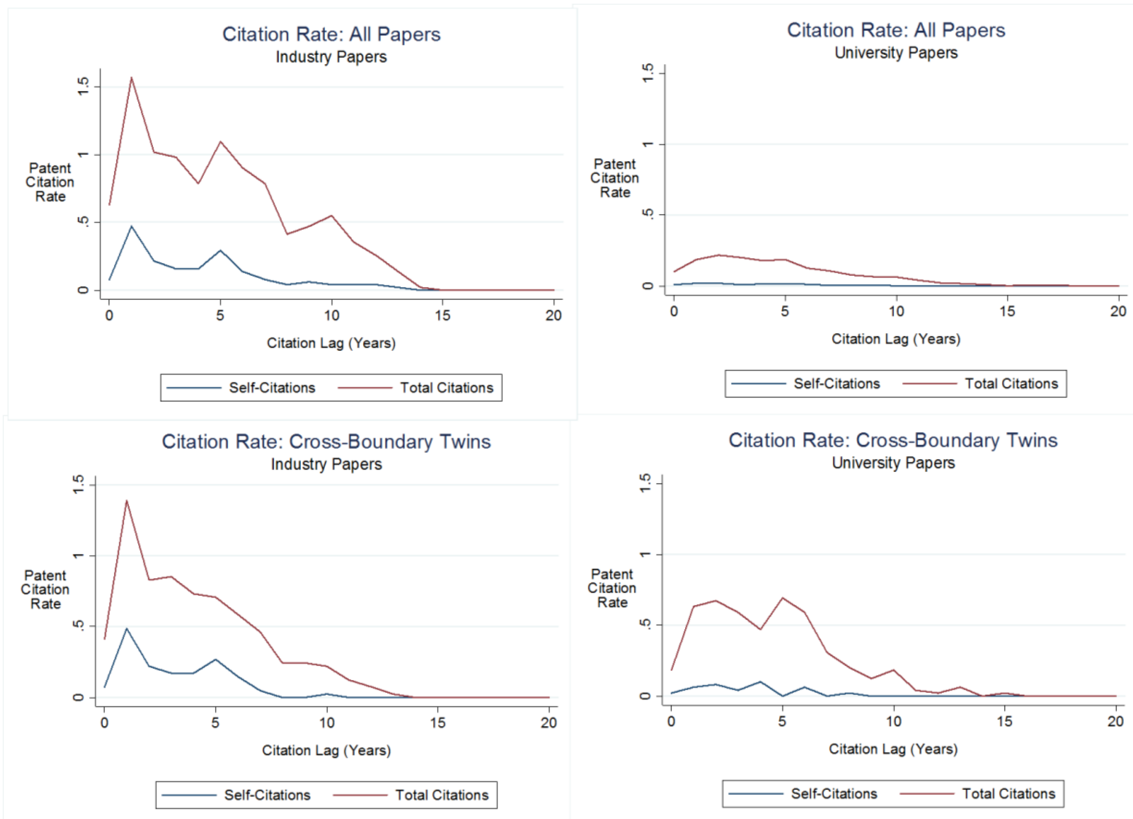

FIGURE 3. Descriptive statistics: Discoverer invention in firms and universities

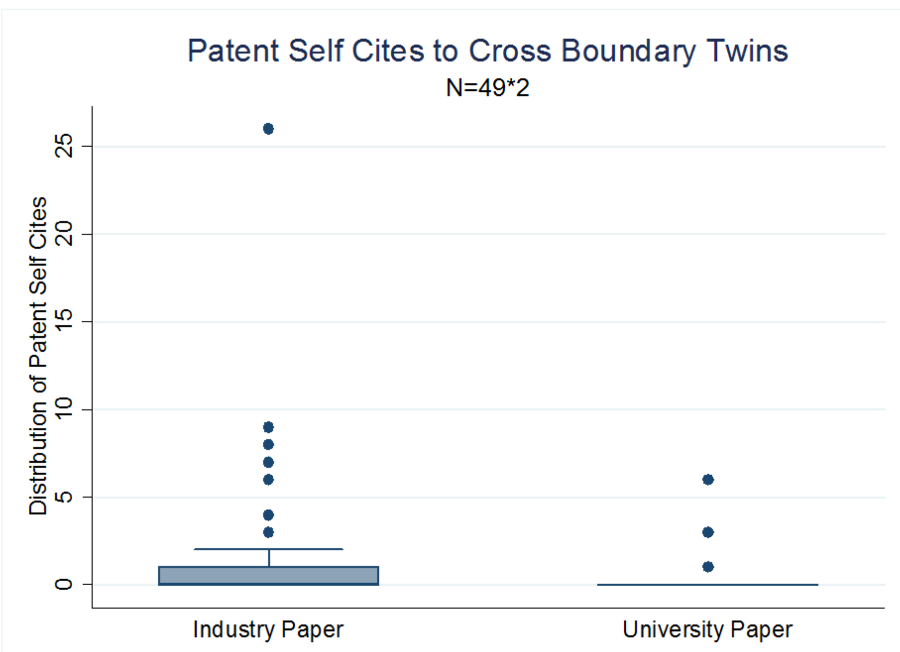

FIGURE 4. Descriptive statistics: Drawing on firm vs. academic knowledge

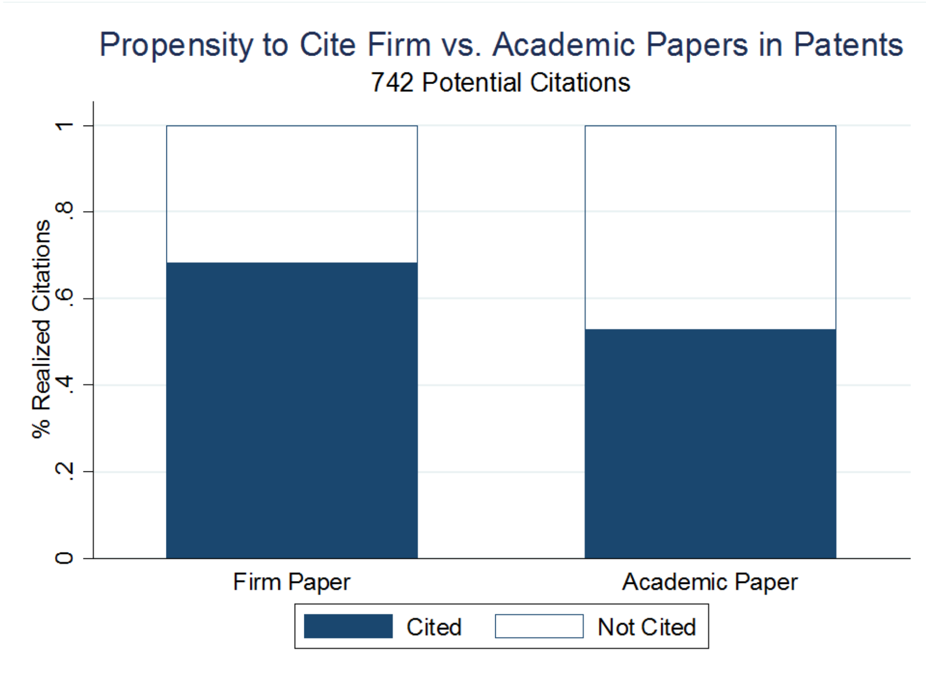

FIGURE 5. Impact of geographic proximity on the rate of paper citation (Predicted Values)