When Collaboration Bridges Institutions:

The Impact of Industry Collaboration on Academic Productivity[1]

Michaël Bikard	Keyvan Vakili	Florenta Teodoridis
London Business School	London Business School	USC Marshall School of Business
mbikard@london.edu	kvakili@london.edu	teodorid@marshall.usc.edu

September 2017

Abstract

Prior research suggests that academic scientists who collaborate with firms may experience lower publication rates in their collaborative lines of work due to industry's insistence on IP protection through patenting or secrecy. The main empirical challenge of examining the effect of industry collaboration on scientific productivity is that research projects that involve industry collaborators may be qualitatively different from those that do not. Hence, any difference in subsequent output of academic scientists who collaborate with industry may be driven by differences in the nature of research projects that attract industry collaborators. To address this issue, we exploit the occurrence of simultaneous discoveries where multiple scientists make roughly the same discovery around the same time. Following a simultaneous discovery, we compare the follow-on research output of academic scientists who collaborated with industry on the discovery with that of academic scientists who did not. We find that academic scientists who collaborated with industry produce more follow-on publications and fewer follow-on patents on their collaborative research lines than their academic peers who did not collaborate with industry. Our results suggest that research lines with both scientific and commercial potential provide an opportunity for a productive division of tasks between academic scientists and their industry counterparts, where the former focus on exploiting the scientific opportunities and the latter focus on the commercial ones. We also show that these effects are particularly salient when the industry partner is an established company rather than a startup.

Introduction

Frequent collaboration between academia and industry is a hallmark of the knowledge economy. These collaborations are often encouraged at the national level,[2] and many studies have emphasized their strategic

[1] All authors contributed equally.

[2] See http://www.nih.gov/news-events/news-releases/nih-fund-collaborations-industry-identify-new-uses-existing-compounds and http://www.nih.gov/news-events/news-releases/nih-launches-collaborative-program-industry-

importance for firms seeking to gain new knowledge, forge new relationships, and yield higher R&D productivity (Cockburn and Henderson 1998; Zucker, Darby, and Armstrong 2002; Owen-Smith and Powell 2004; Lacetera 2009). However, the opposite impact, that of industry collaboration on academic science, remains less understood and more contentious (e.g., Murray 2010; Perkmann et al. 2013). The academic institutional environment highly differs from that of industry (Merton 1973; Dasgupta and David 1994), and some scholars have raised concerns that the corporate emphasis on commercialization might erode academic productivity and weaken the norms of science in academia (Blumenthal, Campbell, et al. 1996; Mowery et al. 2001; Perkmann and Walsh 2009).

At its core, this concern rests on the view that industry's institutional environment conflicts with that of academia. Specifically, firms' emphasis on commercialization and appropriation of their intellectual property rights through patenting and secrecy can limit their academic collaborators' contribution to open science (Blumenthal, Campbell, et al. 1996; Louis et al. 2001; Perkmann and Walsh 2009; Toole and Czarnitzki 2010; Evans 2010), induce a guarded behavior among scientists (Blumenthal, Campbell, et al. 1996; Campbell et al. 2000, 2002), and eventually reduce public disclosure of research output in terms of scientific publications (Czarnitzki, Grimpe, and Toole 2014; Lee 2000; Thursby and Thursby 2002). The adverse effects of industry collaboration on academic scientists' productivity and output could be particularly heightened in areas with higher commercial potential and thus higher appropriation risks. The problem is compounded by concerns of shifts in academic institutional norms from freedom of science (Merton 1973) towards knowledge transfer through commercialization (e.g.,Dasgupta and David 1994; Azoulay, Ding, and Stuart 2009). Prior research shows that industrial practices of secrecy, patenting, and entrepreneurship have already spread across the halls of academia to some extent (e.g., Argyres and Liebeskind 1998; Stuart and Ding 2006; Bercovitz and Feldman 2008).

In contrast, we propose that, under certain circumstances, industry collaboration may actually reinforce traditional characteristics of the academic institutional environment by increasing publication and

researchers-spur-therapeutic-development as examples of funding schemes to boost collaboration between academia and industry.

lowering patenting rates. In research lines that have both scientific and commercial potential, collaboration between academia and industry might lead to a more productive division of tasks and responsibilities, where each side would be able to focus on what their institutional environment rewards most: academia on more fundamental scientific insights, and industry on commercialization (Merton 1973; Aghion, Dewatripont, and Stein 2008; Sauermann and Stephan 2013).

Prior research suggests that scientists sort into either academia or industry according to their preferences. Scientists with a stronger "taste for science" are more likely to join academia; scientists with relatively stronger preference for applied research and monetary incentives sort into industry (Stern 2004; Roach and Sauermann 2010; Agarwal and Ohyama 2013). In addition, the academic institutional environment is particularly amenable to the exchange of knowledge and materials, as well as to curiosity-driven research (Bush 1945; Nelson 1959; Merton 1973; Aghion, Dewatripont, and Stein 2008; Sauermann and Stephan 2013). For all these reasons, academic institutions possess a relative competitive advantage in investigating more basic, fundamental questions, whereas firms maintain a relative advantage in development and commercialization. Therefore, from firms' point of view, collaborating with academia enables them to focus on areas in which they are the strongest, while relying on their academic partners to explore the projects' more fundamental aspects. Active collaboration during the project can ensure that the knowledge created by the academic side will quickly and seamlessly transfer to the corporate partner. Likewise, collaborating with industry can give academic scientists access to resources, skills, and equipment that might be valuable for fulfilling the scientific potential of a line of research, without the pressure to spend much time on its commercialization. The differences between academia and industry in their approaches to scientific research may also enable academic scientists to discover new research opportunities within the domains of their collaborative line of research (Agrawal and Henderson 2002; D'Este and Perkmann 2011). In a sense, then, collaboration might be an arena in which the specificities of the academic institutional environment can be leveraged and reinforced.

Hence, by offering the opportunity for a more productive allocation of tasks on lines of research with both scientific and commercial potential, industry collaboration might bolster academic contribution

to open science while helping firms achieve higher R&D efficiency and better access to academic research. Thus, we expect academic scientists who collaborate with industry in projects that offer both scientific and commercial potential to experience higher subsequent publication rate within the collaborative lines of works. We further expect to see a better division of labor when academic scientists collaborate with established companies, rather than startups, given that the former usually possess better commercialization capabilities and generally more resources to leverage during the collaboration.

The empirical challenge in testing these arguments is considerable because collaboration is a choice. Firms and academic scientists both strategically select the research questions they want to pursue and whether or not they want to pursue them through collaboration. Thus, the type of research projects pursued through university-industry collaboration is likely to be systematically different from those solely pursued in academia. This makes comparisons between academic productivity and engagement with commercialization across different types of projects ill-suited for testing our predictions. Since academic scientists are on average more likely to work on applied projects when they collaborate with firms than when they do not, a finding that academic researchers are on average less scientifically productive when they work with firms might simply indicate that academic scientists work with firms on lines of research that are on average less scientifically promising.

To address this challenge, our empirical strategy holds the line of research constant by exploiting a phenomenon known as multiple or simultaneous discoveries (Merton 1957; Cozzens 1989). Simultaneous discoveries are scientific achievements that are published at the same time by two or more different scientific teams. When two teams make roughly the same discovery at about the same time, they face the same level of scientific and commercialization opportunities following that discovery. In this paper, we use a sample of 33 such simultaneous discoveries, each involving 2 or 3 teams. In each of those instances, at least one academic team collaborated with a firm, and at least one other did not.

For each team disclosing a simultaneous discovery, we track the corresponding author's follow-on contribution to open science and engagement with commercialization for the line of research associated with the simultaneous discovery. We use scientific publications as indicators of contribution to open science

and patents as indicators of engagement with the commercialization process. Specifically, we count the corresponding authors' publications and patents that were built upon the simultaneous discovery – i.e., follow-on publications and follow-on patents respectively. We choose to focus on corresponding authors because these scientists are generally the most senior individuals who take responsibility for the research. Our data crosses different scientific domains, and hence different publication norms, yet the norm for corresponding authors is uniform in our sample.[3] We choose to focus on follow-on publication and patents within the line of research of the simultaneous discovery because, based on prior research, the negative effect of collaboration with industry is rooted in industry's insistence on protecting the project's commercial potential. Note also that firms' involvement in academic science can take different forms. We do not expect to see a publication boost in cases where academic scientists are simply funded by firms. For many academic researchers, industry is just one possible source of funding and many others exist. For a productive division of labor to occur, active collaboration is likely to be needed. We identify such collaborations by tracking academic researchers' co-authorship with scientists working in industry.

Consistent with our predictions, our results suggest that following a simultaneous discovery, academic scientists with industry collaborators produce more follow-on publications and fewer follow-on patents than academic scientists who made the same discovery without an industry collaborator. In our data, academic scientists who collaborate with firms produce 50% more citation-weighted publications and 99% fewer citation-weighted patents than their academic peers who did not collaborate with industry. Moreover, we find that the results are driven by discoveries that combine both high scientific and high commercial potential (as opposed to high scientific potential only) and discoveries where the industry partner is an established firm (as opposed to a start-up). We further show that the higher levels of publication output built upon discoveries with industry partners do not come at the expense of a decline in publication output outside the collaborative research line. Importantly, the effect is only visible in active collaborations

[3] To verify this statement, we contacted several authors on these papers, and they shared the same sentiment. Further, the approach is in line with previous literature (e.g., Furman and Stern, 2011; Furman, Jensen and Murray, 2012; Zucker et al., 1998).

involving co-authorship, and it is not visible in the case of simple sponsorship of academic research by industry grants. The results are robust to alternative specifications and various robustness checks.

Our paper makes three contributions. First, we present evidence that collaboration with industry might increase academic contribution to open science under some circumstances, namely when the project has both scientific and commercial potential. In such cases, industry collaboration appears to sharpen the particularities of the academic institutional environment rather than corrupting it. Our results add more nuance to the idea that academic and industrial environments can be complementary rather than conflicting (Sauermann and Stephan 2013; Rosenberg 1990; Cohen and Levinthal 1990; Hicks 1995; Gittelman and Kogut 2003; Stern 2004; Ding 2011) by highlighting important contingencies—types of project and firm—for a productive collaboration between the two institutions.

Second, our results highlight how collaboration shapes institutions. Even though the increasingly close relationship between academia and industry might jeopardize specific aspects of the academic environment, we present evidence that, under certain conditions, close proximity through collaboration might reinforce institutional differences rather than dilute them.

Finally, this paper is the first to use simultaneous discoveries to explore how team composition shapes the work of academic scientists. We build on the proposition that simultaneous discoveries can be used as an empirical tool to conduct what amounts to "twin studies of scientific knowledge" (Bikard 2012). Because they constitute instances in which the same or very similar discoveries are made around the same time by different teams, this method provides potentially rich insights on the elements that shape research trajectories.

Industry Collaboration and Academic Productivity

Industry collaboration negatively influences academic productivity

Historically, the key mandate of universities has been to generate and disseminate knowledge (Merton 1973). A defining characteristic of academic researchers is autonomy, such that scientists are free to select the projects they undertake and have creative control over the research methods they pursue and how they choose to disseminate their findings (Aghion, Dewatripont, and Stein 2008). Prior research shows that

scientists value this freedom and will accept lower wages to maintain it (Stern 2004). With the passage of Bayh-Dole act of 1980 in the United States, which encouraged universities to more actively engage in patenting and licensing their research outcomes, and with the increased collaboration between industry and academia, many have expressed concerns that too much involvement with industry's commercialization activities and profit-seeking agendas may distort the institutional norms that have long governed academic research (Mowery et al. 2001). In particular, scholars have argued that collaboration between industry and academia acts as a bridge between the two institutional environments, and hence, industrial orientation toward economic gains and secrecy might travel over that bridge and erode the traditional academic emphasis on openness and publication.

Scholars discussing the negative implications of university–industry collaboration point out that industry collaboration is likely to harm academic productivity because industrial practices conflict in many ways with academic values and norms (Merton 1973; Dasgupta and David 1994; Stephan 1996). Industry's emphasis on secrecy might decrease academic productivity because firms may impose publication restrictions on their academic collaborators within the collaborative lines of research (e.g., Lee 2000; Thursby and Thursby 2002). Surveys of academic researchers have associated collaboration with industry with higher levels of secrecy, less sharing in general, and more publication delays (Blumenthal, Causino, et al. 1996; Blumenthal et al. 1997; Louis et al. 2001; Evans 2010; Shibayama, Walsh, and Baba 2012; Czarnitzki, Grimpe, and Toole 2014). For example, in a survey of 210 life-science companies, Blumenthal et al. (1996) found that among the companies that supported academic research, 82% asked academic scientists to keep their findings confidential for some time to allow for patent applications, and that 56% require confidentiality for longer than the time needed for patent application. Relationships with industry might deter sharing because of the emphasis on confidentiality, but the impact might also be deeper, as sharing norms might progressively shift from generalized to more direct forms of exchange (Shibayama, Walsh, and Baba 2012). Similar problems were reported in cases of industry sponsorship: an indirect form of university–industry collaboration. For example, in a recent survey of 1,060 German academic researchers, Czarnitzki, Grimpe, and Toole (2014) found that industry sponsorship increases the expected

probability of publication delay from 14% to 33%, and that of secrecy (complete or partial ban on publishing research results) from 11.2% to 35%. Other research shows that patenting a scientific discovery may negatively affect scientists' investment in follow-on research based on that discovery (Williams, 2013). Overall, then, academic scientists working with industry are likely to find it more difficult to control publication of their research results.

In addition to the issue of publication restrictions, industry collaboration could also harm academic productivity by inducing scientists to undertake commercialization-related activities that are unlikely to lead to scientific publications (Goldfarb 2008; Perkmann and Walsh 2009; Toole and Czarnitzki 2010). Some empirical studies have linked academic scientists' involvement with industry to a decrease in their contribution to open science. In particular, Toole and Czarnitzki (2010) study the impact on academic productivity of life-scientists' involvement in for-profit firms through the SBIR program. Using a case-cohort sampling design, they find that the yearly publication rate of academic scientists decreases by 19% after they become involved in the program, which indicates a considerable "brain drain" away from academic science and toward commercial activities.

The potential negative impact of industry collaboration on academic scientists' contribution to open science seems especially alarming considering that the scientists who work with firms tend to be some of the most productive in their fields (Blumenthal, Campbell, et al. 1996; Zucker and Darby 1996; Stuart and Ding 2006; Toole and Czarnitzki 2010).

Industry collaboration positively influences academic productivity

By examining collaboration as a conduit through which industrial practices flow into academia, prior literature has uncovered an important negative facet of cross-institutional collaboration. In this paper, however, we argue that under certain circumstances, academia–industry collaboration may foster a more productive collaboration and generate higher academic productivity. Collaboration, we argue, is not just a channel through which practices and norms can diffuse across institutions; rather, it is also an arena in which the complementarities across institutional environments can be leveraged given the right circumstances. Specifically, we argue that in projects with both scientific (basic) and commercial (applied)

opportunities, academia and industry can mutually benefit if each party focuses on what it does best; i.e., if academia focuses on more fundamental aspects, and industry on more applied aspects. In such projects, industry collaboration might reinforce the particularities of the academic institutional environment by creating an opportunity for academic scientists to devote their efforts mostly towards exploiting the scientific potential of the project while being minimally distracted by the commercialization process.

A large stream of research argues that firms can benefit from collaborating with academia on research projects. Academic publications can improve R&D efficiency if that scientific knowledge is relevant to the firm (Nelson 1982; Cohen, Nelson, and Walsh 2002; Fleming and Sorenson 2004). Academic research can also give firms new market opportunities (Toole and Czarnitzki 2007; Zucker, Darby, and Brewer 1998) and help them overcome technological barriers (Cohen, Nelson, and Walsh 2002; Furman and MacGarvie 2009). However, for academic scientists to have the incentive to invest in a research project, they should be able to reap the returns of their investment in terms of academic publications and scientific credit. Contributing to open science through publication has tremendous value to scientists in academia because it allows them to take part in the scientific "cycles of credit" (Latour and Woolgar 1986). For academic scientists, publishing is not only a norm, it is also an essential driver of scientific careers. Stern (2004) shows that scientists are willing to accept lower wages if given more academic freedom in their jobs. Hence, firms may benefit from letting their academic partners focus on their scientific research; this would be particularly true when the project at hand requires considerable basic and fundamental research (Lacetera 2009). Past research shows that scientists with higher preference for basic research are more likely to join universities (Agarwal and Ohyama 2013). Hence, partnering with university scientists can help firms complement their competitive advantage in applied research and commercialization with the basic research capabilities of their academic collaborators.

Academic scientists can likewise benefit from collaborating with industry. First, the presence of an industry partner can help academic scientists focus on their basic research activities and leave the commercialization aspects, such as patenting and licensing, to their industry collaborators. Past research suggests that academia is relatively less efficient than industry when it comes to commercialization (XXX).

Hence, we expect academic scientists with industry collaborators to be able to spend more time pursuing the scientific potential of their collaborative line of research, since these researchers benefit from the commercialization capabilities of their industry partners.

Second, industry can be a source of valuable skills, equipment, and material and financial resources (Mansfield 1995; D'Este and Perkmann 2011; Tartari and Breschi 2012). D'Este and Perkmann (2011), for example, surveyed 1,088 grant holders from the UK's Engineering and Physical Sciences Research Council who had engaged with industry and asked them about their motivation for doing so. Interestingly, their results highlight four key drivers: access to funding, commercialization, learning, and access to in-kind resources. In the latter category, the respondents appeared to value in particular access to material, followed by access to research expertise and access to equipment.

Third, industry can be a source of new ideas within a particular line of research (Mansfield 1995; Murray 2002; D'Este and Perkmann 2011). Industry and academia usually differ in how they approach similar research questions; they use different perspectives to formulate research questions, and different methods to investigate them (Rosenberg 1994; Mansfield 1995; Siegel et al. 1999). Hence, collaboration with industry can give academic scientists access to new and different insights and knowledge assets, which could in turn increase their research output. A long stream of research on innovation suggests that diversity in perspective, insights, and knowledge could help innovative teams break away from intellectual lock-in and explore new, otherwise unexplored territories (Jeppesen and Lakhani 2010; Audretsch, Dohse, and Niebuhr 2009; Phene, Fladmoe-Lindquist, and Marsh 2006; Rosenkopf and Nerkar 2001; Gavetti and Levinthal 2000; March 1991). Using a theoretical model, Hong and Page (2004) demonstrate how a diverse team of randomly selected individuals may outperform a homogeneous team of best-performing ones because individuals with varied backgrounds and perspectives could evaluate the same opportunities differently, and hence exploit a more heterogeneous set of opportunities. Furthermore, academic scientists' access to alternative sets of resources in industry may open up new opportunities for investigating a phenomenon.

Fourth, academic scientists can benefit from industry collaboration since past research suggests that, beyond caring about academic credit, scientists also care about the ultimate application of their research and its potential use outside academia (Murray 2010; D'Este and Perkmann 2011; Stephan 2012). To this end, an industry partner can give academic scientists additional incentive to invest in a line of research, without requiring them to directly engage with the applied aspects of their research. In other words, the mere presence of an industrial partner might afford academic scientists the confidence that their research results will have more chances to be used outside of academia, and hence acts as an additional incentive mechanism to further invest in the collaborative line of research.

Finally, collaboration with industry can help scientists invest more heavily in lines of research with commercialization opportunities. Prior research suggests that industry involvement in a research domain, particularly in the form of patenting, may dissuade academic scientists from investing in follow-on research due to potential litigation risks (Heller and Eisenberg 1998; Murray and Stern 2007). This might lead to an under-exploitation of such research lines, which could give firms incentives to correct the expected deficit through collaboration. In other words, academic scientists who have industry collaborators may face lower hindrance in these research domains.

Following these arguments, we postulate that scientists collaborating with industry on certain projects might therefore generate higher levels of follow-on contribution to open science (e.g., publication) and lower involvement in commercialization activities (e.g., patenting). Industry collaboration in a particular line of research might accentuate some institutional characteristics that are distinctively academic. We therefore predict the following:

Hypothesis 1 (H1): For a given research project with both academic and applied potential, an academic scientist with an industry collaborator is likely to produce more follow-on scientific publications than an academic scientist without an industry collaborator.

Hypothesis 2 (H2): For a given research project with both academic and applied potential, an academic scientist with an industry collaborator is likely to produce fewer follow-on patents than an academic scientist without an industry collaborator.

An efficient and productive allocation of tasks and responsibilities between academia and industry along their domains of expertise requires that the collaborative project at hand offer high potential for both scientific research and commercialization, which enables each party to focus on what it does best. To offer additional insight, we focus on the moderating effect of the commercial potential of a line of research.[4]

The above predictions challenge the idea that the pursuit of economic returns in industry negatively affects the research output of collaborating academic scientists. If, as prior research suggests, appropriation concerns lead to more secrecy, one should expect secrecy and IP protection to play a more significant role in projects with more commercial potential. In contrast, our arguments suggest that higher commercial potential may provide better conditions for a productive allocation of tasks and efforts. Hence, we predict the following:

> *Hypothesis 3 (H3): Commercial potential of a research project positively moderates the relationship between academia–industry collaboration and the follow-on scientific publications of academic scientists.*

Furthermore, we expect the experience level of industry partners to positively moderate the benefits that academic scientists accrue from their collaboration with industry. Established companies, as opposed to startups, have deeper industry-specific knowledge, better commercial capabilities, and generally more resources to share with their academic partners. Given the central role of these factors in facilitating a productive division of tasks and responsibilities between the two types of institutions, we expect that academic scientists who collaborate with established companies yield more benefit than those who collaborate with startups. This prediction contradicts what prior research suggests. Established companies are usually more embedded in the industry mindset of rent-seeking and protecting intellectual property rights, and hence are expected to be more cautious with publishing the outcomes of their research projects. Startups, on the other hand, may use these publications as signals of their quality and be more supportive

[4] While we believe scientific potential as a moderating factor could also offer interesting deeper insights, we are limited by our empirical context, in which all research projects have considerable scientific potential. The lowest number of academic citations to a paper in our sample is 66, which already puts that paper in the top 1% of highly cited papers in its respective field.

of contribution to open science. However, we argue that the knowledge assets and commercialization capabilities of established companies better support the hypothesized productive collaboration between academia and industry. Hence, we predict the following:

> *Hypothesis 4 (H4): Collaboration with established companies positively moderates the relationship between academia–industry collaboration and the follow-on scientific publications of academic scientists.*

Empirical Strategy

The core challenge in examining the impact of industry–academia collaboration on subsequent research productivity of academic scientists is that scientists strategically select their research partners and the research questions they want to pursue. Thus, the research projects that are pursued through university–industry collaboration may fundamentally differ from those pursued solely in academia or industry. Consequently, any estimated effect of university–industry collaboration on subsequent research productivity captures both the effect of collaboration and the effect of choosing a particular research question.

In an ideal setting, we would observe the same research question assigned randomly and simultaneously to two teams: one composed only of academic collaborators and one that includes industry collaborators. We would then want to observe follow-on publication and patenting efforts of academic scientists from both teams within the respective line of research and to measure the differences between those efforts. In the absence of such an ideal experiment, we use a novel empirical approach exploiting the existence of simultaneous discoveries.

Simultaneous discoveries are scientific discoveries that are independently conducted by two or more teams of scientists and reported in what we call paper twins: papers published on the same discovery around the same time. Paper twins are hence dual instances of the same scientific discovery by different teams in different environments. The following example resulted from a discovery that was simultaneously made by scientists from the Harvard Medical School and scientists at the University of Utah and Heidelberg University working in collaboration with Myriad Pharmaceutical:

Strack et al. (September 2003) "AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding" ***Cell***

(Affiliations: Department of Cancer Immunology and AIDS Dana-Farber Cancer Institute and Department of Pathology, Harvard Medical School)

"HIV-1 and other retroviruses exit infected cells by budding from the plasma membrane, a process requiring membrane fission. The primary late assembly (L) domain in the p6 region of HIV-1 Gag mediates the detachment of the virion by recruiting host Tsg101, a component of the class E vacuolar protein sorting (Vps) machinery. We now show that HIV Gag p6 contains a second region involved in L domain function that binds AIP1, a homolog of the yeast class E Vps protein Bro1. Further, AIP1 interacts with Tsg101 and homologs of a subunit of the yeast class E Vps protein complex ESCRT-III. AIP1 also binds to the L domain in EIAV p9, and this binding correlates perfectly with L domain function. These observations identify AIP1 as a component of the viral budding machinery, which serves to link a distinct region in the L domain of HIV-1 p6 and EIAV p9 to ESCRT-III."

von Schwedler et al. (September 2003) "The protein network of HIV budding" ***Cell***

(Affiliations: Department of Biochemistry, University of Utah; Department of Pathology, Heidelberg University; Myriad Pharmaceuticals Inc.)

"HIV release requires Tsg101, a cellular factor that sorts proteins into vesicles that bud into multivesicular bodies (MVB). To test whether other proteins involved in MVB biogenesis (the class E proteins) also participate in HIV release, we identified 22 candidate human class E proteins. These proteins were connected into a coherent network by 43 different protein-protein interactions, with AIP1 playing a key role in linking complexes that act early (Tsg101/ESCRT-I) and late (CHMP4/ESCRT-III) in the pathway. AIP1 also binds the HIV-1 p6 and EIAV p9 proteins, indicating that it can function directly in virus budding. Human class E proteins were found in HIV-1 particles, and dominant-negative mutants of late-acting human class E proteins arrested HIV-1 budding through plasma and endosomal membranes. These studies define a protein network required for human MVB biogenesis and indicate that the entire network participate in the release of HIV and probably many other viruses."

These excerpts describe two sets of independent findings regarding the role of the same protein components in the HIV viral budding process. One of the papers (Strack et al. 2003) was written by a research team within academia, the other (von Schwedler et al. 2003) in a collaborative effort between

scientists in academia and a firm (Myriad Pharmaceuticals Inc.). The two papers were published back-to-back in the September 2003 volume of the journal *Cell*.

The use of simultaneous discoveries helps us control for the strategic selection into particular research questions by different scientists and hence isolate the effect of university–industry collaboration. In particular, we assume that when two different teams report the same discovery, they face the same level of follow-on scientific and commercial opportunity to pursue based on that discovery. In other words, the simultaneous discovery opens up the same level of scientific and commercial opportunity for both teams of scientists around the same time. This approach allows us to observe academic scientists' behavior with respect to subsequent knowledge production based on the discovery when influenced by industry collaboration, while controlling for the otherwise hard-to-observe counterfactual of the same project undertaken absent industry collaboration.

We use a sample of paper twins in which one paper is authored by only academic scientists, and the other results from collaboration between academia and industry. The main (corresponding) author in both samples is an academic scientist.[5] An important characteristic of the studies in our sample is that they attract the attention of both academia and industry. Hence, if one thinks of the landscape of scientific publications in terms of Donald Stokes' (1997) typology, twin papers would most likely be better representatives of papers in the Pasteur's quadrant where discoveries have both scientific and commercial potential. This selection corresponds directly to the boundary conditions of our theoretical argument. At the same time, it suggests that our empirical findings do not necessarily extend to publications that do not have this characteristic. We elaborate more on the generalizability of our findings in the discussion section.

For each simultaneous discovery, we compare the follow-on research output of the main authors after the publication of the simultaneous discovery. We define follow-on research output of a scientist as

[5] The corresponding author is clearly identified on each publication as the main point of contact between the authorship team and all external parties. The corresponding author is generally the more senior individual who takes responsibility for the research, provides the intellectual input, and approves the research design and protocols. In about 70% of the papers in our sample, the corresponding author is the same as the last author. In most cases where the corresponding author is the first author, we could confirm that the first author was the senior scientist running the lab.

the count of scientific papers that list the scientist as an author and cite her initial paper reporting the simultaneous discovery. By fixing the initial discovery, we can attribute differences in the follow-on research output of scientists to the composition of their teams for their initial discovery, i.e., the difference in their involvement with industry. To better isolate the effect of university–industry collaboration, we further control for several observable characteristics of the main scientists and their respective teams. Table 1 lists the set of variables used in our analyses and explains how they are defined and constructed.

-- Table 1 approximately here –

Formally, we use the following regression equation to estimate the effect of university–industry collaboration on follow-on publications of corresponding scientists on each twin paper:

$$follow_on_publication_count_{ip} = f(\beta academia_industry_collaboration_{ip} + \alpha X_{ip} + discovery_p + \varepsilon_{ip})$$

where $follow_on_publication_count_{ip}$ is the number of academic publications of scientist i citing the initial simultaneous discovery p after its publication. X_{ip} is a vector of controls at the scientist-discovery level, including scientist's academic experience, her past academic publication and patent stock, a dummy capturing whether the scientist was affiliated with an institution in the United States, the number of authors listed on the paper reporting the initial simultaneous discovery p, the share of authors affiliated with institutions in the United States, the share of female authors, and a dummy capturing whether the paper was supported by any industry grants or not. Simultaneous-discovery fixed effects are captured by $discovery_p$. The inclusion of discovery fixed effects ensures that all the comparisons are between the main scientists that made a simultaneous discovery around the same time and hence faced the same subsequent scientific and commercial opportunity. In other words, we do not rely on comparisons across discoveries and only compare two scientists who made the same discovery around the same time. Because the dependent variable is a count, we use a Poisson model with robust standard errors. The results are robust to the use of a more restricted set of control variables as well as to alternative specifications including negative binomial and OLS estimation models.

Following our theoretical arguments, to the extent that collaboration between academia and industry leads to a better allocation of tasks and responsibilities, and the associated positive effect on research output of academic scientists overcomes the potential negative effects proposed in prior literature, we expect β to be positive and significant. Otherwise, in the absence of any positive effect, or where the negative effects argued in prior literature are greater than the positive effects argued in this study, β would be negative or non-significant.

We use the same regression method to identify the impact of academia–industry collaboration on follow-on patenting by each scientist (H2). The follow-on patenting efforts are measured as the count of patents that cite the original discovery (in their non-patent references) and list the main scientist behind the discovery as an inventor. Note that the follow-on patents are not necessarily patents granted on the original discovery itself. Rather, these are generally patents filed on subsequent commercial developments informed by the simultaneous discovery. Following our theoretical arguments, we expect the effect of university–industry collaboration to be negative.

To test hypothesis H3 and H4, we add interaction terms between our dummy variable $academia_industry_collaboration_{ip}$ and variables capturing whether the discovery has high commercial potential or not (H3) and whether the industry partner on discovery p was an established company or not (H4). We use the total count of patents that have cited any of the twin papers reporting a simultaneous discovery as a proxy for the commercial potential of that discovery. Given that the measure is the same for all twin papers reporting the same discovery, we do not include the variable alone in the regressions, since its effect is captured by discovery fixed effects. We then categorize discoveries that are above median in commercial potential as discoveries with high commercial potential. The results are robust to using a continuous measure of commercial potential. To identify whether a company partner was established or not, we use the age of the company partner at the time of each discovery. Companies with more than 10 years of experience at the time of collaboration are categorized as established. Again, we only include the variable in its interaction form, because its direct effect is captured by the discovery fixed effect.

The results are robust to neighboring cutoff years for this categorization. We also perform several additional robustness analyses that we discuss in more detail below.

Data

The data for this study are based on the first automatically and systematically collected dataset of simultaneous discoveries. The dataset contains 72 scientific publications disclosing 33 simultaneous discoveries.[6] As noted above, the main authors on all publications are affiliated with academia, but the collaborating teams differ in industry involvement. Specifically, within each pair of papers reporting the same discovery, one publication lists authors from both industry and academia, and the other lists authors solely from academia. 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 scientific literature, not only in the same papers, but also within the same parentheses, or adjacently (Cozzens 1989). Figure 1 summarizes the algorithm used to construct the sample of simultaneous discoveries. The algorithm is based on the method developed by Bikard (2012).

-- Figure 1 approximately here –

We collected data about each paper from three bibliographical databases: ISI Web of Science, Scopus, and PubMed. Details about the main (corresponding) scientists were collected from various public sources, such as the scientists' public profiles. Follow-on publications (through October 2013) were collected from two bibliographical databases: Scopus and PubMed. Most discoveries in our sample are in life sciences in areas such as cell biology, stem cell, genetics, and immunology. The sample includes discoveries in physics as well.

Our follow-on patent data are the intersection of two sets of patents which we built separately. First, we collected all the patents that build on each simultaneous discovery. To do so, we built a script that used the Harvard Dataverse patent data (Li et al. 2014), which is based on USPTO data, to identify the set of patents that cite either of the papers disclosing the simultaneous discovery. Patent citations to papers

[6] Twenty-seven of the simultaneous discoveries were reported in two publications, and 6 were disclosed in three papers.

indicate that the inventors drew on the knowledge disclosed in the scientific paper as part of the invention process (e.g., Azoulay, Graff Zivin, and Sampat 2012; Roach and Cohen 2012). Therefore, we consider those patents to constitute inventive activity building on the simultaneous discovery. Second, we collected all the US patents awarded by each corresponding author in our sample. To do so, we searched the USPTO website for each scientist's name. To address the issue of different inventors having the same name (disambiguation), we assessed each patent separately and established the inventor's identity based on application date, inventor's address, and patent topic. Our dataset of follow-on patents by the academic scientists is the intersection between these two sets.

We provide summary statistics for our main variables in Table 2. The average number of follow-on publications for each author-paper was 20.9, and the average number of follow-on patents was 0.44. The oldest of the simultaneous discoveries dated back to 1996, and the most recent occurred in 2008. On average, about 10 authors collaborated on each discovery paper. In 64% of discovery papers, the corresponding author was affiliated with a US institution. Also, 61% of the authors on each discovery paper were affiliated with US institutions. A typical main author had about 16 years of experience and about 113 papers and 2 patents before the publication of the discovery paper. Approximately 30% of authors on the discovery papers were women, and about 26% of the discovery papers were supported by industry grants.

-- Table 2 approximately here --

The list of all universities and non-for-profit research organizations involved with the discoveries in our sample is relatively long. Of the 750 scientists involved in these simultaneous discoveries, about 77 researchers are affiliated with a company and the rest with an academic institution. In total, there are 183 unique academic or non-for-profit research institutions in our sample. Some of the most frequent institutions are the Institute for Cancer Research (London, UK), Harvard University, the University of California, the University of Michigan, Yale University, the Salk Institute, the University of Texas, Dana-Farber Cancer Institute, Max Planck Institute, the University of Iowa, the University of Utah, Columbia University, Duke University, Imperial College London, Scripps Research Institute, and Rockefeller University.

Furthermore, in total, 25 unique companies are involved as collaborators on the papers in our sample. A few companies are involved with more than one discovery in our sample. The established companies in our sample are Novartis, Merck, Schering-Plough, Bayer, GSK, Genentech, Xerox Corporation, the EMMES Corporation, Mitsubishi, Elan Corporation, and Amgen. Of these, the youngest is Amgen, which was established in 1980. The list of startups includes Millennium Pharmaceuticals, NeuroSpheres Ltd., Athersys, Virologic, Caprion Pharmaceuticals, Ligand Pharmaceuticals, Lexicon Pharmaceuticals, Regeneron Pharmaceuticals, Perlegen Sciences, Sangamo Biosciences, ATABIS GmbH, deCODE Genetics, and Myriad Pharmaceuticals. The startups in our sample were all founded after 1991.

Results

As a first step, we report some simple within-twin comparisons. In approximately 59% of the cases, the academic authors who had industry collaborators produced more follow-on research than those who had no industry collaborators. Also, when comparing the twins, after excluding the cases where none of the corresponding authors of either twin produced any follow-on patents, in 64% of the cases the academic scientists who had an industry collaborator produced fewer follow-on patents than the academic scientists who had no industry collaborators on their twin discovery. The comparative statistics are in line with Hypotheses 1 and 2. Next, we provide the results of our regression analysis.

In Table 3, we report results of our main estimating equation: the impact of university–industry collaboration on the count of main scientists' follow-on academic publications and patents. Model (3-1) shows the impact of university–industry collaboration on the count of follow-on academic publications. Consistent with Hypothesis 1, the results suggest that collaboration with industry increases the follow-on research output of academic scientists by about 37%, which amounts to approximately 7.7 additional follow-on papers published based on the original discovery—an increase of 0.44 standard deviations. In model (3-2), we adjust our dependent variable to take into account the quality of follow-on academic publications. As in prior research, we use the number of citations that an academic publication received as a proxy for its quality. The results show that industry collaboration has a similar positive effect on the

quality-adjusted number of follow-on academic publications. Industry collaboration is associated with an additional 1,542 citation-weighted follow-on publications—an increase of 0.35 standard deviations.

-- Table 3 approximately here --

Model (3-3) shows the estimated impact of university–industry collaboration on follow-on patenting by the main scientists. The estimates suggest that, compared with those having no industry collaborators, scientists who collaborated with industry partners on their simultaneous discoveries produce more than 99% fewer patents building on their discovery. In model (3-4), we again adjust the number of patents based on their quality, as measured by the number of citations. The estimated effect of industry collaboration is similar to that reported in model (3-3). The results are consistent with Hypothesis 2. The large estimates suggest that the effect is largely driven by whether the corresponding authors patent or not. Hence, we repeated our estimations using Logit and Probit estimation models with a dummy dependent variable indicating whether the main author has any follow-on patents (DV=1) or not (DV=0) based on her simultaneous discovery.[7] The results are qualitatively similar to those reported in Table 3. The standard deviations are larger potentially due to the loss of approximately 20% of the observable variance from the cases where both corresponding authors of matched twin papers have produced follow-on patents based on their simultaneous discovery. Overall, our finding on the impact on industry collaboration on follow-on patenting persists. Nevertheless, given our small sample size and the rare nature of patenting incidents, the estimations are rather sensitive to the choice of statistical model.

In model (4-1) of Table 4, we investigate how the commercial potential of a discovery moderates the effect of industry collaboration on scientists' follow-on research output. The estimate indicates that that the positive effect of academia–industry collaboration on scientists' research productivity is driven by the subset of discoveries with high commercial potential. For the category of discoveries with high commercial

[7] We could not obtain convergence on our estimations based on Logit and Probit models with the full set of control variables due to the limited observable variance and few degrees of freedom. Hence, in our robustness checks with Logit and Probit estimations, we use a more limited set of control variables. More specifically, we only control for whether the main author is affiliated with an institution in the United States (particularly because we use the sample of patents filed with the USPTO) and the main author's past patenting experience.

potential, collaboration with industry increases scientists' follow-on publications by more than 54%, calculated based on the net of direct effect and the interaction effect. The estimated effect is equivalent to 11 additional follow-on publications—an increase of approximately 0.65 standard deviations. In contrast, for discoveries with low commercial potential, the effect is negative and not statistically significant. The negative effect suggests that academia–industry collaboration on projects with low commercial potential may lead to negative effects in research productivity of academic scientists, as argued in prior research. In other words, in the absence of the conditions for a productive and complementary allocation of tasks and responsibilities, collaboration with industry may potentially harm academic scientists' research output. Model (4-2) repeats the same estimation using the quality-adjusted number of follow-on publications as the dependent variable. The estimates are similar and overall consistent with our prediction in Hypothesis 3.

-- Table 4 approximately here –

Next, we test our fourth hypothesis; namely, the moderating role of collaboration with established companies versus startups, and present estimation results in models (4-3) and (4-4) of Table 4. Consistent with our prediction in Hypothesis 4, the estimated interaction effect in model (4-3) indicates that collaboration with an established company increases scientists' follow-on research output by approximately 67% (equivalent to approximately 14 additional follow-on publications or an increase of 0.8 standard deviations), whereas collaboration with startups (companies with less than 10 years of experience) has no significant effect on scientists' research output. Model (4-4) repeats the previous estimation with a quality-adjusted number of follow-on publications as the dependent variable. The results are consistent and support Hypothesis 4.

Alternative Explanations and Robustness Tests

While we control for each scientist's past publication stock in all models reported in Table 3, one may still be concerned that the positive impact of academia–industry collaboration on scientists' follow-on research output may be driven by the selection of more productive scientists into collaboration with industry. In other words, it is possible that scientists with higher research productivity are more likely to collaborate with industry, and that the same scientists continue to publish more based on the simultaneous discoveries.

To address this concern, we test the relationship between the likelihood of collaboration with industry and each author's past research productivity, patenting activity, academic experience, and location. Since the dependent variable is a 0/1 dummy variable, we use a logit estimation model. We also include discovery fixed effects to ensure that the comparison takes place between scientists reporting the same discovery. The estimates in model (5-1) of Table 5 suggest that collaboration with industry is negatively associated with past research productivity. Further, the estimates indicate that collaboration with industry is not correlated with past patenting activity of scientists or with their academic experience. In model (5-2), we repeat the estimations using an OLS regression and find similar results. Overall, the estimates suggest that the positive effects reported in Table 3 (in support of H1) are not driven by selection of more-productive scientists into collaboration with industry.

-- Table 5 approximately here --

We also compare the non-twin, academia-only papers that were published in the same year as a twin paper by each corresponding author with industry collaborators to non-twin papers that were published by the corresponding author on the other twin paper without industry collaborators. A higher similarity in non-twin papers by the corresponding authors behind a simultaneous discovery provides further evidence that the differences in follow-on research output of scientists who collaborated with industry is not driven by scientists' average higher quality. To compare the non-twin papers, we use a regression analysis, estimating the difference in several characteristics of non-twin papers on whether the main author had industry collaborators on her twin paper. We compare the non-twin papers on three main dimensions: the number of academic citations received, whether they received any patent citations subsequently, and the number of authors on each paper. The results are reported in Table 6. The estimates show no significant difference in the non-twin, academia-only papers produced by the matched main authors. We also find no significant difference in the number of non-twin papers produced by the corresponding authors who had industry collaborators and those who did not have any industry collaborators, in the year of their twin publication. Overall, the results support the assumption that the corresponding authors who had industry collaborators had similar output at the time of publishing their twin papers.

-- Table 6 approximately here --

Even though the previous analysis shows that the differences in the experience of the corresponding authors on twin papers is unlikely to explain the differences in their follow-on research output, it is still possible that the corresponding authors with industry collaborators had a more senior or more experienced team of co-authors on their side. Hence, the increase in their follow-on research output may be associated with the presence of more experienced co-authors and not that of an industry partner. To address this concern, we first identified the status of every author on every paper at the time of its publication using publicly available information. We then categorized the authors into research faculty, university research staff (such as graduate students and postdocs), industry scientists, and unidentified. We could identify the status of 60% of the authors in our sample. We then checked whether there is any significant difference in the status composition of teams between twin papers. Models 1 to 3 of Table A1 in the appendix show the results of estimating the difference in number of faculty, the number of research staff (grad students and postdocs), and the number of individuals with unidentified status between the teams with industry collaborators and the teams without. Furthermore, we collected all pre-discovery publications by all authors in each authorship team using the Scopus database. To the extent that the team with industry collaboration has more experienced individuals, we should expect that team to also have a higher level of total pre-discovery publications. Model 4 of Table A1 shows the results of estimating the authorship team's total number of pre-discovery publications on whether the team involved industry partners. We find no significant difference between the two teams.

Yet another possible alternative interpretation of our results is that the increase in scientists' follow-on research output due to industry collaboration may have come at the expense of decline in research output in areas outside the discovery's line of research. In other words, what we are interpreting as an increase in research output driven by a complementary allocation of tasks may simply be due to a shift in the research direction of scientists collaborating with industry. In Table 7, we test the impact of academia–industry collaboration on scientists' research output outside the discovery line of research. The effect on both the simple count and the quality-adjusted count of publications, reported in models (7-1) and (7-2) respectively,

shows that we find no evidence that industry collaboration affects scientists' research output in areas that are not directly related to the collaborative line of research.

-- Table 7 approximately here –

Furthermore, to ensure that our results are driven not by industry sponsorship but by active collaboration with industry, we also test the effect of industry grants on follow-on research productivity of scientists using a larger sample of 1,236 papers disclosing 578 simultaneous discoveries. Most discoveries in this larger sample were reported by teams of only academic scientists. However, some teams received industry sponsorship for their discovery, and some did not. To build the sample of industry grants, we manually collected the acknowledgements sections of all the papers in the sample and extracted all mentions of funding support from industry. Our results in Table 8 suggests that industry sponsorship has no significant effect on follow-on research productivity of scientists, and that the effects reported here rely on the mechanism of active collaboration between academia and industry, as hypothesized.

-- Table 8 approximately here –

Finally, in Table A2 in the appendix we repeat the estimations reported in interaction models (4-2) and (4-4) using Negative Binomial and OLS regressions with discovery fixed effects. The results are similar to those reported with the Poisson model. While the estimated negative effect of industry collaboration on projects with low commercial potential is larger in these estimates, the net positive effect on citation-weighted follow-on publications based on discoveries with high potential (calculated as the net of the direct effect and the interaction effect) is comparable to that reported in Table 4. Our results are also robust to the inclusion of a more restricted set of control variables (to allow for more degrees of freedom) as shown in Table A3 in the appendix.

Discussion and Conclusion

Our paper draws upon literatures on collaboration, institutions, and scientific human capital to examine the implications of industry–academia collaborations for academic research. Specifically, we model the influence of industry collaboration on academic scientists' follow-on publishing and patenting rates. This question has received increasing attention in light of widespread concerns that industry collaboration might

corrupt traditional academic norms. In this study we argue that when a research line has both scientific and commercial potential, the collaboration between academia and industry can support better allocation of tasks and responsibilities between the two types of institutions, leading to greater complementarity through collaboration. More specifically, academic scientists can achieve higher levels of scientific output by gaining access to skills, resources, and equipment from firms, while leaving the applied and commercial aspects of research to their industry partner. Industry collaborators can also be a source of new insights and expertise for academic scientists.

The empirical challenge in investigating this question is considerable, because academic scientists do not randomly sort into industry collaboration. In other words, prior findings that industry collaboration is associated with fewer publications and more patents in academia might be a consequence of norm contamination, but they might also reflect the fact that academic scientists collaborate with firms on their more applied projects with lower scientific potential. This paper addresses this challenge by focusing on a set of simultaneous discoveries, in which an academic scientist collaborating only with other academic scientists makes the same discovery as one collaborating with industry. We observe 33 such events, which we use as a small set of natural experiments to investigate the existence of systematic differences in the subsequent publishing and patenting patterns of scientists with and without industry collaborators for those lines of work.

Holding the line of research constant, we find that academic scientists who collaborate with industry produce significantly more follow-on publications and fewer patents than academic scientists who do not collaborate with industry. The estimated effects are relatively large. Based on the same line of research, collaboration with industry is associated with 36% more publications and 99% fewer patents. Moreover, the effects do not appear to be driven either by the sorting of more capable scientists into collaborations with industry, or by the fact that firms might induce academic scientists to invest more in their line of research of interest at the expense of other research projects. Rather, we find evidence consistent with the existence of a better allocation of tasks and responsibilities between academia and industry, wherein academic scientists focus more on publishing and invest less in patenting. In line with this

argument, our main effects are most pronounced for the lines of research that have high commercial potential and those projects that involve collaborations with mature, established companies.

Our conversations with several authors of simultaneous discoveries in our sample further corroborate our quantitative estimations. For example, when asked about the role of university and industry in their collaborative projects on drug development, a leading faculty member at a prominent West Coast university noted:

> More and more, you are seeing the enlightened companies getting their molecules out there into academic labs so people can play around with them. [Companies] make drugs and reagents, but they often don't have the very sophisticated and elegant systems that we have to interrogate things in vivo…We have these fairly elegant, genetically-engineered models that really recapitulate what happens in human cancer.

Another academic scientist from a prominent university made a similar remark, highlighting that industry collaboration can provide access to capabilities that can speed up research:

> The fact that we collaborated with [firm] meant that we could go faster. That did give us a competitive advantage because they had this capability of doing high throughput two-hybrid screening.

Scientists in industry shared similar sentiments regarding the complementarities between academia and industry in their collaborative projects. As an example, when asked about his simultaneous discovery and the allocation of tasks and efforts between his team in industry and his collaborators in academia, a senior scientist in a large East Coast pharmaceutical company responded:

> [In] this case we were sharing our mice with the academic labs… [and] they would do the analysis of the mice. The company would set up a material transfer agreement that handles the legalities of sending the mice over to the academic labs. There would be hypotheses that would be set up front in terms of the experiments that would be carried on in the academic lab. Then we'll send the mice, they would do the experimental stuff and then they looked to see if they find something interesting.

> The company cannot do all the science itself. So one of the things I did at [my company] was to make reagents and send them to the academic lab as soon as possible, so you maximize the value of the reagents that you are generating. So it's a way to connect with academic labs, a way to know the science before it comes out as a publication, it's a way of helping the field move forward.

When asked whether the publication of findings would help the company's competitors, he responded that

> …making new medicine has a 99% failure rate. So if you're going to be super-conservative and be closed and siloed away from everyone, you are actually increasing the odds that you are not going to be able to deliver anything. So I think what you gain is more than what you lose by being transparent.

Overall, our results emphasize a hitherto under-recognized positive aspect of the academic institutional environment for collaboration with industry. Prior literature has highlighted that academia can be a valuable source of knowledge for firms, but that its traditional emphasis on freedom and openness clashes with industry's focus on secrecy and intellectual property (Blumenthal, Campbell, et al. 1996; Louis et al. 2001; Perkmann and Walsh 2009; Toole and Czarnitzki 2010; Evans 2010). Our findings complement this stream of work by indicating that the differences in the way the value of scientific knowledge is captured in academia and in industry present an opportunity for a complementary allocation of tasks and responsibilities for lines of research with both scientific and commercial potential. Thus, the distinctiveness of the academic environment can be reinforced—rather than jeopardized—by industry collaboration.

Note that our theory and empirical analysis are not generalizable to the whole spectrum of scientific projects but instead apply to those projects with both scientific and commercial potential. In other words, it is not advisable based on our findings to attract an industry collaborator on every scientific project. In our setting, one should keep two kinds of selection in mind. First, the discoveries in our sample are particularly important. They attracted the attention of several teams of scientists, and journal editors collectively decided more than once that the discovery was worthy of publication. Moreover, our method of identifying simultaneous discoveries is based on co-citation. For these reasons, we only observe relatively highly cited

discoveries. Second, the twins in our sample are based on discoveries that have attracted the interest of industry. Hence, if one thinks of the landscape of scientific publications in terms of Donald Stokes' (1997) typology, twin papers would be better representatives of papers in the Pasteur's quadrant where discoveries have both scientific and commercial potential.

How large is the population of papers to which our findings are potentially generalizable? While it is difficult to assess how many papers, of the massive universe of scientific publications, have both scientific and commercial potential, we can show that the sample is telling relative to the full stock of publications of authors in our sample. Specifically, we collected all publications by the corresponding authors in our sample that were published in the same year as the twin papers. We used the same publication year to denote twins to ensure that the difference between the twin and non-twin publications by each scientist is not driven by her tenure or changes in the opportunity landscape in her field over time. A main author in our sample has produced on average 9.1 other papers (st.dev=8.6) in the same year as her twin paper. Next, we categorized all papers published by the corresponding authors into four categories: 1) twin papers without industry collaborators; 2) twin papers with industry collaborators; 3) non-twin, academia-only papers by scientists who had no industry collaborators on their twin papers; and 4) non-twin, academia-only papers by scientists who had industry collaborators on their twin papers. We compared these four categories on two main dimensions: the number of received academic citations, and whether the papers were cited in any subsequent patent (0/1 dummy). We used citations in subsequent patents as a proxy for the commercial potential of a scientific publication. We further log-normalized the number of academic citations to address their skewed distribution. Figures 2 and 3 compare the distributions of these two variables across the four categories. There are a few notable observations. First, the twin papers have on average higher rates of both paper citations and patent citations. This is expected considering the nature of twin discoveries. The distribution of paper citations as well as the existence of a patent citation is tilted towards the right in both figures. However, there is a good overlap between the distributions of twin papers and non-twin, academia-only papers. Specifically, Figure 4 overlays the distributions for twins without industry collaborators (first

category) onto the distributions for academia-only, non-twins (second and fourth categories). The overlap suggests that twin papers still represent an important part of academia-only publications.

-- Figures 2, 3 and 4 approximately here –

Furthermore, we break the non-twin papers down into four groups: 1) those with above-mean paper citations and at least one patent citation, 2) those with below-mean paper citations and at least one patent citation, 3) those with above-mean paper citations and no patent citations, and 4) those with below-mean paper citations and no patent citations. We used the logged paper citations for the whole sample to calculate the mean. Based on this categorization, 39% of non-twin, academia-only papers by scientists who had industry collaborators on their twin discovery fall into the first group (high scientific and commercial potential). Similarly, 31% of non-twin, academia-only papers by scientists who did not have industry collaborators on their twin discovery also fall into the first group. If we use a more stringent criterion to define papers with high scientific potential (selecting only papers that have more citations than the average twin-paper minus one standard deviation), still about 11% of non-twin papers fall into high scientific potential and high commercial potential category. Overall, our analysis suggests that while twin papers certainly do not represent all academic papers, they are not outliers without any resemblance to other academic papers.

Our findings are not without limitations and therefore open the door to further research. Our empirical strategy relies on using simultaneous discoveries to ensure that the compared scientists are working on the same line of research. Still, industry collaboration may not be randomly assigned in our data; hence there may still be some systematic differences between the scientists who work with firms and those who do not. To mitigate this concern, we control for the observable characteristics that may contour scientists' research behavior (e.g., past productivity, experience, gender), highlighted in prior literature, and we run a number of robustness tests. Yet we cannot completely rule out that our results may be in part driven by unobservable differences across individuals.

We use insights from the collaboration literature to cast new light on debates about the implications of cross-institutional differences. Our findings constitute a first step in highlighting some of the

circumstances under which institutional diversity is a source of opportunity for organizing work in more productive ways. In situations where the same work presents differential values across disparate institutional environments, collaboration might make it possible to take advantage of cross-institutional differences. Such cross-institutional collaborations can be found in various other settings that involve public–private partnerships. Our study calls for future research on contingencies and factors under which, instead of creating unproductive conflicts of interests, misallocation of resources and efforts, and wasteful organizational resistance, such collaborations can increase productivity for all involved parties.

References

Agarwal, Rajshree, and Atsushi Ohyama. 2013. "Industry or Academia, Basic or Applied? Career Choices and Earnings Trajectories of Scientists." *Management Science* 59 (4): 950–70. doi:10.1287/mnsc.1120.1582.

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): 617–35.

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

Argyres, Nicholas S., and Julia Porter Liebeskind. 1998. "Privatizing the Intellectual Commons: Universities and the Commercialization of Biotechnology." *Journal of Economic Behavior & Organization* 35 (4): 427–54. doi:10.1016/S0167-2681(98)00049-3.

Audretsch, David, Dirk Dohse, and Annekatrin Niebuhr. 2009. "Cultural Diversity and Entrepreneurship: A Regional Analysis for Germany." *The Annals of Regional Science* 45 (1): 55–85. doi:10.1007/s00168-009-0291-x.

Azoulay, Pierre, Waverly W. Ding, and Toby Stuart. 2009. "The Impact of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output." *The Journal of Industrial Economics* 57 (4): 637--676.

Bercovitz, Janet, and Maryann Feldman. 2008. "Academic Entrepreneurs: Organizational Change at the Individual Level." *Organization Science* 19 (January): 69–89. doi:10.1287/orsc.1070.0295.

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

Blumenthal, David, Eric G. Campbell, Anderson MS, Causino N, and Louis K. 1997. "Withholding Research Results in Academic Life Science: Evidence from a National Survey of Faculty." *JAMA* 277 (15): 1224–28. doi:10.1001/jama.1997.03540390054035.

Blumenthal, David, Eric G. Campbell, Nancyanne Causino, and Karen Seashore Louis. 1996. "Participation of Life-Science Faculty in Research Relationships with Industry." *N Engl J Med* 335 (23): 1734–39. doi:10.1056/NEJM199612053352305.

Blumenthal, David, Nancyanne Causino, Eric G. Campbell, and Karen Seashore Louis. 1996. "Relationships between Academic Institutions and Industry in the Life Sciences — An Industry Survey." *New England Journal of Medicine* 334 (6): 368–74. doi:10.1056/NEJM199602083340606.

Bush, Vannevar. 1945. "Science: The Endless Frontier." *Transactions of the Kansas Academy of Science (1903-)* 48 (3): 231–64.

Campbell, Eric G., B. R. Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N. A. Holtzman, and D. Blumenthal. 2002. "Data Withholding in Academic Genetics: Evidence From a National Survey." *JAMA* 287 (4): 473.

Campbell, Eric G, Joel S Weissman, Nancyanne Causino, and David Blumenthal. 2000. "Data Withholding in Academic Medicine: Characteristics of Faculty Denied Access to Research Results and Biomaterials." *Research Policy* 29 (2): 303–12. doi:10.1016/S0048-7333(99)00068-2.

Cockburn, Iain, and Rebecca Henderson. 1998. "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery." *The Journal of Industrial Economics* 46 (2): 157–82.

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.

Czarnitzki, Dirk, Christoph Grimpe, and Andrew A. Toole. 2014. "Delay and Secrecy: Does Industry Sponsorship Jeopardize Disclosure of Academic Research?" *Industrial and Corporate Change*, May, dtu011. doi:10.1093/icc/dtu011.

Dasgupta, Partha, and Paul A. David. 1994. "Toward a New Economics of Science." *Research Policy* 23 (5): 487–521.

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. doi:10.1007/s10961-010-9153-z.

Evans, James A. 2010. "Industry Collaboration, Scientific Sharing, and the Dissemination of Knowledge." *Social Studies of Science* 40 (5): 757–91. doi:10.1177/0306312710379931.

Fleming, Lee, and Olav Sorenson. 2004. "Science as a Map in Technological Search." *Strategic Management Journal* 25: 909–28.

Furman, Jeffrey, and Megan MacGarvie. 2009. "Academic Collaboration and Organizational Innovation: The Development of Research Capabilities in the US Pharmaceutical Industry, 1927–1946*." *Industrial and Corporate Change* 18 (5): 929–61. doi:10.1093/icc/dtp035.

Gavetti, Giovanni, and Daniel Levinthal. 2000. "Looking Forward and Looking Backward: Cognitive and Experiential Search." *Administrative Science Quarterly* 45 (1): 113–37. doi:Article.

Goldfarb, Brent. 2008. "The Effect of Government Contracting on Academic Research: Does the Source of Funding Affect Scientific Output?" *Research Policy* 37 (1): 41–58. doi:10.1016/j.respol.2007.07.011.

Heller, Michael A., and Rebecca S. Eisenberg. 1998. "Can Patents Deter Innovation? The Anticommons in Biomedical Research." *Science* 280 (5364): 698–701. doi:10.1126/science.280.5364.698.

Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences of the United States of America* 101 (46): 16385–89. doi:10.1073/pnas.0403723101.

Jeppesen, Lars Bo, and Karim R. Lakhani. 2010. "Marginality and Problem-Solving Effectiveness in Broadcast Search." *Organization Science* 21 (5): 1016–33. doi:10.1287/orsc.1090.0491.

Lacetera, Nicola. 2009. "Different Missions and Commitment Power in R&D Organizations: Theory and Evidence on Industry-University Alliances." *Organization Science* 20 (3): 565–82. doi:10.1287/orsc.1080.0366.

Latour, Bruno, and Steve Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.

Lee, Yong S. 2000. "The Sustainability of University-Industry Research Collaboration: An Empirical Assessment." *The Journal of Technology Transfer* 25 (2): 111–33. doi:10.1023/A:1007895322042.

Louis, Karen Seashore, Lisa M. Jones, Melissa S. Anderson, David Blumenthal, and Eric G. Campbell. 2001. "Entrepreneurship, Secrecy, and Productivity: A Comparison of Clinical and Non-Clinical Life Sciences Faculty." *The Journal of Technology Transfer* 26 (3): 233–45. doi:10.1023/A:1011106006976.

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

March, James G. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* 2 (1): 71–87. doi:10.2307/2634940.

Merton, Robert K. 1957. “Priorities in Scientific Discovery: A Chapter in the Sociology of Science.” *American Sociological Review* 22 (6): 635–59.

———. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Mowery, David C., Richard R. Nelson, Bhaven N. Sampat, and Arvids A. Ziedonis. 2001. “The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980.” *Research Policy* 30 (1): 99–119. doi:10.1016/S0048-7333(99)00100-6.

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. doi:10.1086/653599.

Murray, Fiona, and Scott Stern. 2007. “Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-Commons Hypothesis.” *Journal of Economic Behavior & Organization* 63 (4): 648–87. doi:10.1016/j.jebo.2006.05.017.

Nelson, Richard. 1959. “The Simple Economics of Basic Scientific Research.” *Journal of Political Economy* 67 (3): 297–306.

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

Oettl, Alexander. 2012. “Reconceptualizing Stars: Scientist Helpfulness and Peer Performance.” *Management Science* 58 (6): 1122–40. doi:10.1287/mnsc.1110.1470.

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

Perkmann, Markus, Valentina Tartari, Maureen McKelvey, Erkko Autio, Anders Broström, Pablo D’Este, Riccardo Fini, et al. 2013. “Academic Engagement and Commercialisation: A Review of the Literature on University–industry Relations.” *Research Policy* 42 (2): 423–42.

Perkmann, Markus, and Kathryn Walsh. 2009. “The Two Faces of Collaboration: Impacts of University-Industry Relations on Public Research.” *Industrial and Corporate Change* 18 (6): 1033–106.

Phene, Anupama, Karin Fladmoe-Lindquist, and Laurence Marsh. 2006. “Breakthrough Innovations in the U.S. Biotechnology Industry: The Effects of Technological Space and Geographic Origin.” *Strategic Management Journal* 27 (4): 369–88. doi:10.1002/smj.522.

Roach, Michael, and Henry Sauermann. 2010. “A Taste for Science? PhD Scientists’ Academic Orientation and Self-Selection into Research Careers in Industry.” *Research Policy* 39 (3): 422–34. doi:10.1016/j.respol.2010.01.004.

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

Rosenkopf, Lori, and Atul Nerkar. 2001. “Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry.” *Strategic Management Journal* 22 (4): 287–306.

Sauermann, Henry, and Paula Stephan. 2013. “Conflicting Logics? A Multidimensional View of Industrial and Academic Science.” *Organization Science* 24 (3): 889–909. doi:10.1287/orsc.1120.0769.

Shibayama, Sotaro, John P. Walsh, and Yasunori Baba. 2012. “Academic Entrepreneurship and Exchange of Scientific Resources Material Transfer in Life and Materials Sciences in Japanese Universities.” *American Sociological Review*, August, 0003122412452874. doi:10.1177/0003122412452874.

Siegel, Donald, David Waldman, Jonathan Silberman, Albert Link, C. A. Santa Rosa, and A. S. U. West. 1999. “Assessing the Impact of Organizational Practices on the Performance of University Technology Transfer Offices: Quantitative and Qualitative Evidence.”

Stephan, Paula. 1996. “The Economics of Science.” *Journal of Economic Literature* 34 (3): 1199–1235.
———. 2012. *How Economics Shapes Science*. Cambridge, Massachusetts: Harvard University Press.
Stern, Scott. 2004. “Do Scientists Pay to Be Scientists?” *Management Science* 50 (6): 835–53. doi:10.1287/mnsc.1040.0241.
Stuart, Toby E., and Waverly W. Ding. 2006. “When Do Scientists Become Entrepreneurs? The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences.” *American Journal of Sociology* 112 (1): 97–144. doi:10.1086/502691.
Tartari, Valentina, and Stefano Breschi. 2012. “Set Them Free: Scientists’ Evaluations of the Benefits and Costs of University–industry Research Collaboration.” *Industrial and Corporate Change* 21 (5): 1117–47. doi:10.1093/icc/dts004.
Thursby, Jerry G., and Marie C. Thursby. 2002. “Who Is Selling the Ivory Tower? Sources of Growth in University Licensing.” *Management Science* 48 (1): 90–104.
Toole, Andrew A., and Dirk Czarnitzki. 2007. “Biomedical Academic Entrepreneurship through the SBIR Program.” *Journal of Economic Behavior & Organization*, Academic Science and Entrepreneurship: Dual engines of growth, 63 (4): 716–38. doi:10.1016/j.jebo.2006.05.011.
———. 2010. “Commercializing Science: Is There a University ‘Brain Drain’ from Academic Entrepreneurship?” *Management Science* 56 (9): 1599–1614. doi:10.1287/mnsc.1100.1192.
Zucker, Lynne G., and Michael R. Darby. 1996. “Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry.” *Proceedings of the National Academy of Sciences* 93 (23): 12709–16.
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.

Figure 1- An Automated and Systematic Method to Generate A List Of Simultaneous Discoveries

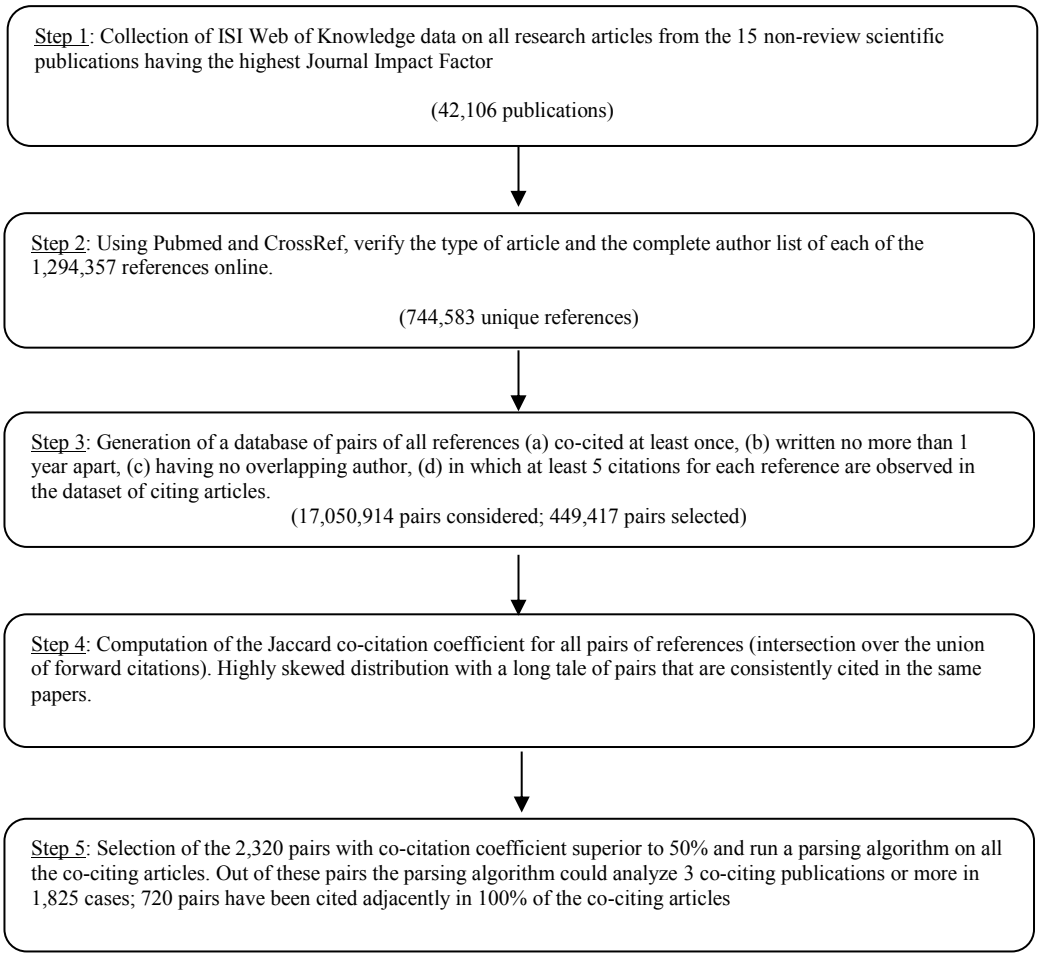
Step 1: Collection of ISI Web of Knowledge data on all research articles from the 15 non-review scientific publications having the highest Journal Impact Factor
(42,106 publications)
Step 2: Using Pubmed and CrossRef, verify the type of article and the complete author list of each of the 1,294,357 references 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 – the academic citation distribution of twin and non-twin papers published by corresponding authors in the same year

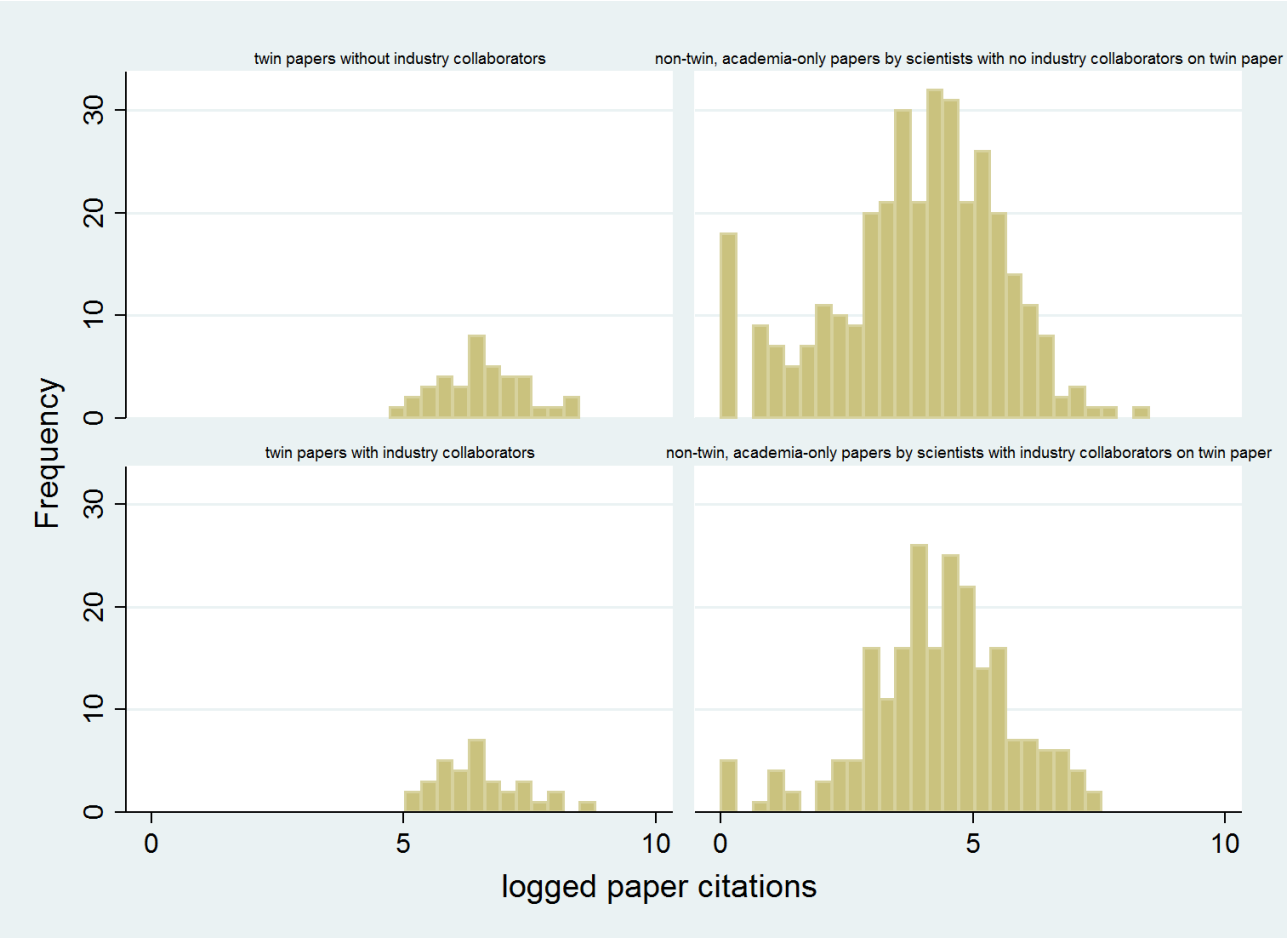

Figure 3 – the patent citation (0/1) distribution of twin and non-twin papers published by corresponding authors in the same year

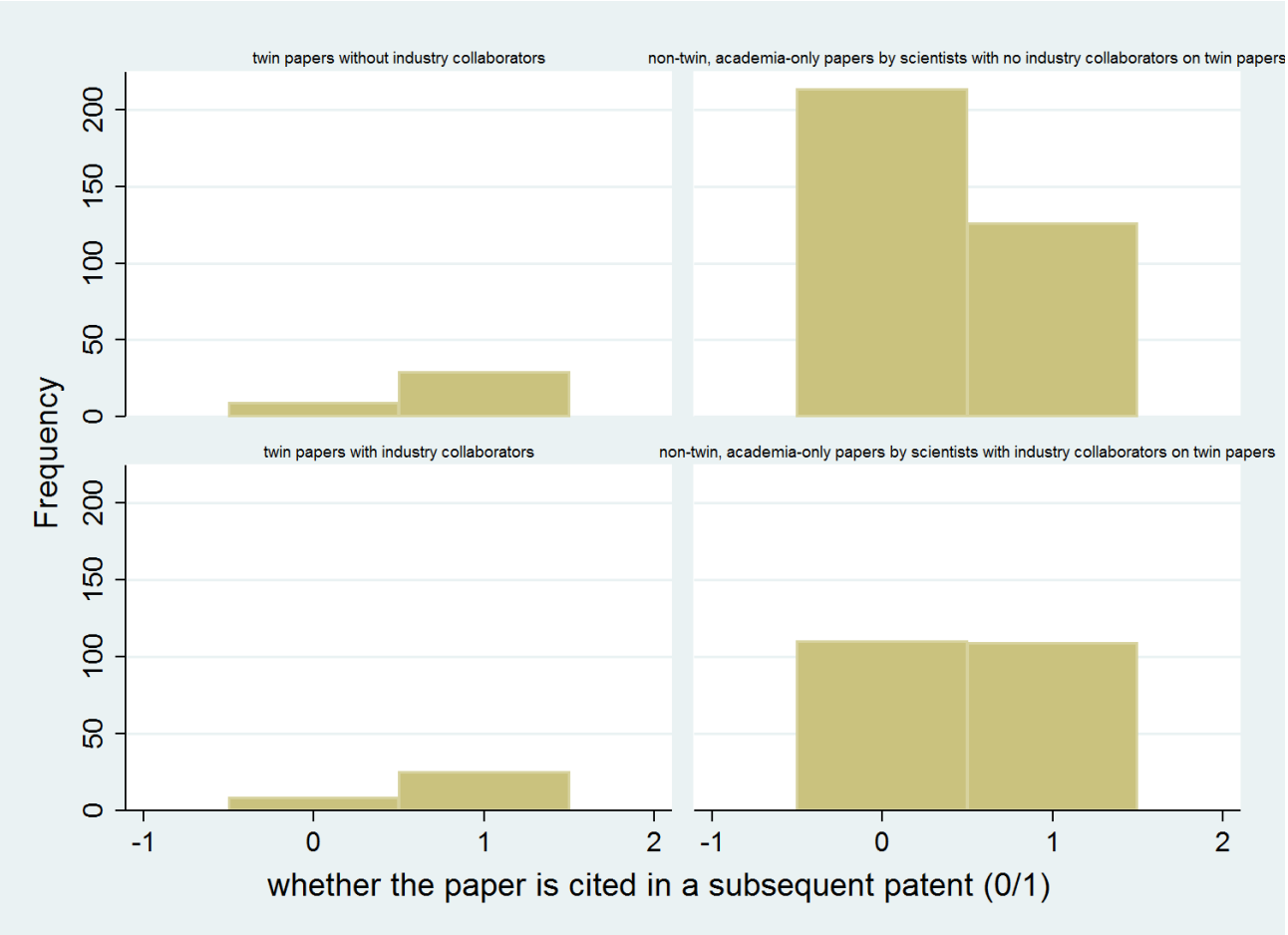

Figure 4– overlap in academic citations between academia-only twins and non-twins by both groups of corresponding authors

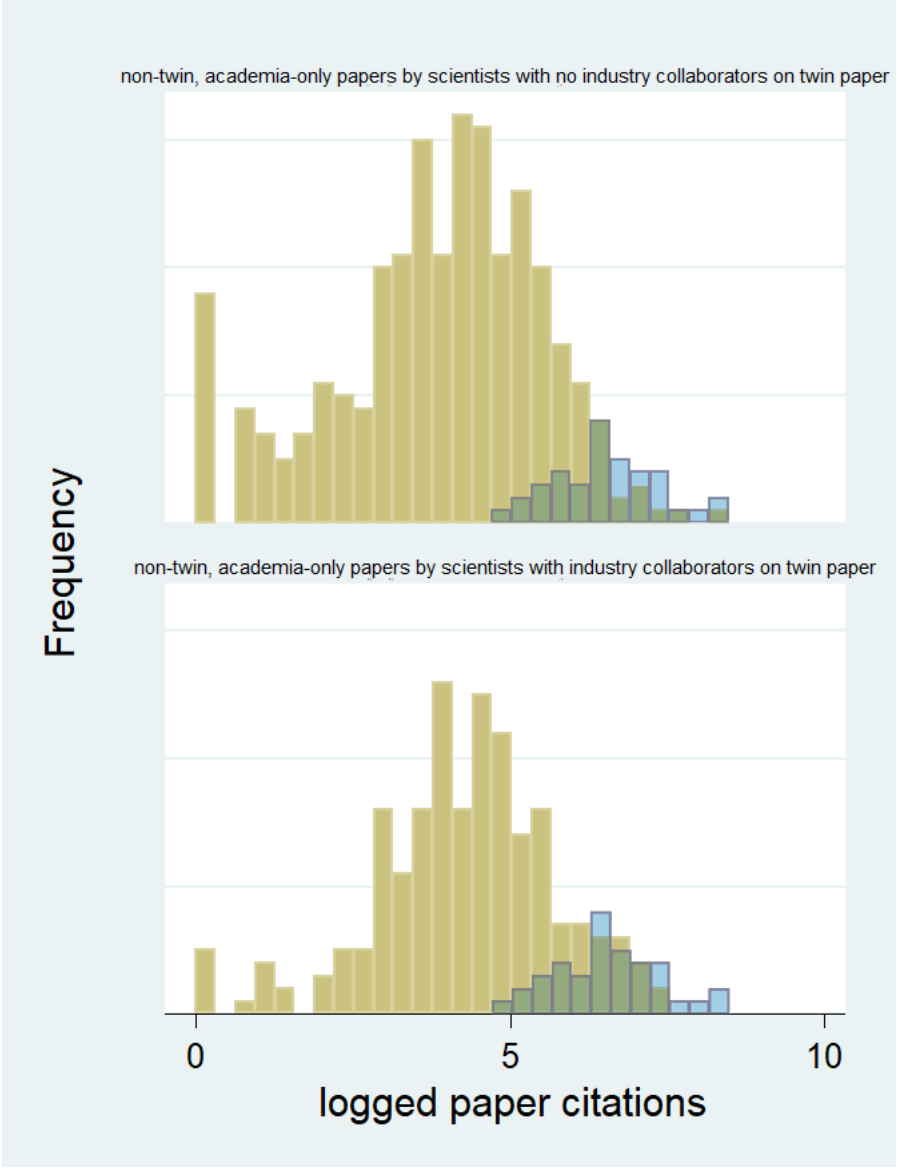

Table 1- Variable Definitions

Variable	Definition
Main author's follow-on papers	Number of papers by the main scientists of a simultaneous discovery citing the original paper that reports the discovery
Main author's citation-weighted follow-on papers	Citation-weighted number of papers by the main scientists of a simultaneous discovery citing the original paper that reports the discovery.
Main author's follow-on patents	Number of patents by the main scientists of a simultaneous discovery citing the original paper that reports the discovery
Main author's citation-weighted follow-on patents	Citation-weighted number of patents by the main scientists of a simultaneous discovery citing the original paper that reports the discovery
Main author's total post-twin papers outside the simultaneous discovery line of research	Total Number of papers by the main scientists of a simultaneous discovery that do not cite the original paper reporting the discovery
Academia-industry collaboration	A dummy variable equal to 1 if the paper that reports a simultaneous discovery is a collaborative work between academia and industry; 0 otherwise.
Number of authors	The number of authors on the paper that reports the simultaneous discovery
Main author's experience	Number of years since author's latest degree
Main author affiliated with a U.S. institution	A dummy variable equal to 1 if the main author on the paper reporting a simultaneous discovery is affiliated with an institution in the U.S.; 0 otherwise
Share of authors affiliated with U.S. institutions	Share of authors affiliated with U.S. institutions on the paper that reports a simultaneous discovery
Share of female authors	Share of female authors on the paper that reports a simultaneous discovery
Industry grant	A dummy variable equal to 1 if the paper reporting a simultaneous discovery is supported by an industry grant, 0 otherwise.
Discovery's commercial potential	Total number of patents that cite a simultaneous discovery

Table 2- Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Max
Main author's follow-on papers	72	20.944	17.489	3	85
Main author's citation-weighted follow-on papers	72	3,041.583	4,414.692	66	26,560
Main author's follow-on patents	72	0.444	0.962	0	4
Main author's citation-weighted follow-on patents	72	0.583	1.392	0	7
Main author's total post-discovery non-twin papers	72	249.903	255.040	5	1365
Academia-industry collaboration	72	0.458	0.502	0	1
Number of authors	72	10.236	9.737	2	65
Main author's past paper count	72	113.153	115.926	1	447
Main author's past patent count	72	1.875	4.587	0	32
Main author's experience	72	16.239	8.496	-2	39
Main author affiliated with a U.S. institution	72	0.639	0.484	0	1
Share of authors affiliated with U.S. institutions	72	0.614	0.408	0	1
Share of female authors	72	0.298	0.185	0	0.667
Industry grant	72	0.264	0.444	0	1
Discovery's commercial potential	72	26.750	62.380	0	345

Table 3- Impact of Academia–industry Collaboration on Follow-On Publications and Patents

	POISSON MODEL			
	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S FOLLOW-ON PATENTS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PATENTS
	(3-1)	(3-2)	(3-3)	(3-4)
Academia–industry collaboration (0/1)	0.313***	0.410***	-7.514*	-6.513**
	(0.118)	(0.126)	(4.051)	(2.954)
Number of authors	-0.018**	0.012	0.692***	0.655***
	(0.012)	(0.011)	(0.264)	(0.216)
Log(main author's past paper count+1)	0.158**	0.551***	-4.083	-3.068**
	(0.078)	(0.074)	(3.806)	(2.746)
Log(main author's past patent count+1)	0.034	-0.037	1.726	1.143
	(0.091)	(0.112)	(1.712)	(1.326)
Log(main author's academic experience+3)	0.323	-0.330	5.457	4.941
	(0.208)	(0.269)	(4.588)	(3.301)
Main author affiliated with a U.S. institution	0.040	1.280***	22.740	19.862
	(0.294)	(0.432)	(23.669)	(18.040)
Share of authors affiliated with U.S. institutions	-0.730**	-1.580***	-12.240	-10.522
	(0.333)	(0.422)	(19.194)	(15.206)
Share of female authors	-0.912	-1.499**	-16.803	-14.991
	(0.566)	(0.673)	(16.792)	(12.603)
Industry grant	-0.265	-0.208	7.723	6.732*
	(0.250)	(0.269)	(5.588)	(4.016)
Twin fixed effects	Yes	Yes	Yes	Yes
Observations	72	72	72	72
Log likelihood	-258.588	-14,593.686	-23.412	-24.043
Paper-twin FE	33	33	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table 4- Role of Commercial Potential and Industry Partner Type

	POISSON MODEL			
	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS
	(4-1)	(4-2)	(4-3)	(4-4)
Academia–industry collaboration (0/1)	-0.198	-0.070	-0.033	-0.051
	(0.261)	(0.279)	(0.178)	(0.213)
Academia–industry collaboration × Discovery has high commercial potential	0.631**	0.529*		
	(0.279)	(0.302)		
Academia–industry collaboration × Industry partner is established			0.549***	0.716***
			(0.214)	(0.276)
Number of authors	-0.017**	0.012	-0.015	0.013
	(0.008)	(0.011)	(0.009)	(0.012)
Log(main author's past paper count+1)	0.150**	0.544***	0.142**	0.513***
	(0.070)	(0.075)	(0.067)	(0.067)
Log(main author's past patent count+1)	0.065	0.007	-0.024	-0.086
	(0.087)	(0.114)	(0.088)	(0.102)
Log(main author's academic experience+3)	0.208	-0.453	0.402	-0.308
	(0.221)	(0.282)	(0.181)	(0.236)
Main author affiliated with a U.S. institution	-0.102	1.189***	-0.067	1.090***
	(0.313)	(0.422)	(0.276)	(0.358)
Share of authors affiliated with U.S. institutions	-0.554*	-1.467***	-0.612**	-1.496***
	(0.312)	(0.401)	(0.292)	(0.363)
Share of female authors	-0.824	-1.430**	-1.185**	-1.806***
	(0.519)	(0.682)	(0.532)	(0.577)
Industry grant	-0.283	-0.253	-0.247	-0.271
	(0.216)	(0.271)	(0.220)	(0.248)
Twin fixed effects	Yes	Yes	Yes	Yes
Observations	72	72	72	72
Log likelihood	-248.811	-14,032.785	-249.623	-12,935.984
Paper-twin FE	33	33	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table 5- Likelihood of Collaborating with Industry based on Main Author's Past Performance

	LOGIT MODEL	OLS MODEL
	DV = ACADEMIA-INDUSTRY COLLABORATION	DV = ACADEMIA-INDUSTRY COLLABORATION
	(5-1)	(5-2)
Log(main author's past paper count+1)	-0.839**	-0.171
	(0.381)	(0.102)
Log(main author's past patent count+1)	0.358	0.069
	(0.456)	(0.147)
Log(main author's academic experience+3)	-0.187	-0.027
	(1.029)	(0.300)
Main author affiliated with a U.S. institution	0.621	0.134
	(0.870)	(0.252)
Industry grant	0.606	0.121
	(0.959)	(0.301)
Twin fixed effects	Yes	Yes
Observations	72	72
R-squared	0.110	0.162
Log likelihood	-42.401	
Paper-twin FE	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table 6- The comparison between non-twin papers of scientists with industry collaborators and those without industry collaborators

	POISSON MODEL	LOGIT MODEL	POISSON MODEL
	DV = NUMBER OF RECEIVED ACADEMIC CITATIONS	DV = WHETHER THE PAPER CITED IN A PATENT (=1, 0 otherwise)	DV = NUMBER OF AUTHORS
	(6-1)	(6-2)	(6-3)
Academia–industry collaboration on the twin published in the same year (0/1)	0.031 (0.154)	0.153 (0.197)	0.102 (0.086)
Twin fixed effects	Yes	Yes	Yes
Observations	447	372	447
Log-likelihood	-45572.015	-217.310	-1515.973

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table 7- Impact of Academia–industry Collaboration on Main Author's Research Direction

	POISSON MODEL	
	DV = MAIN AUTHOR'S TOTAL POST-TWIN PAPERS OUTSIDE THE SIMULTANEOUS DISCOVERY LINE OF RESEARCH	DV = MAIN AUTHOR'S TOTAL CITATION-WEIGHTED POST-TWIN PAPERS OUTSIDE THE SIMULTANEOUS DISCOVERY LINE OF RESEARCH
	(7-1)	(7-2)
Academia–industry collaboration (0/1)	-0.042 (0.096)	0.094 (0.088)
Number of authors	0.011 (0.007)	-0.002 (0.010)
Log(main author's past paper count+1)	0.913*** (0.080)	0.809*** (0.094)
Log(main author's past patent count+1)	-0.086 (0.066)	-0.001 (0.054)
Log(main author's academic experience+3)	-0.429 (0.339)	0.343 (0.236)
Main author affiliated with a U.S. institution	0.683 (0.452)	0.358 (0.556)
Share of authors affiliated with U.S. institutions	-0.755* (0.391)	-0.489 (0.461)
Share of female authors	0.462 (0.338)	0.544 (0.440)
Industry grant	-0.303*** (0.114)	-0.362** (0.153)
Twin fixed effects	Yes	Yes
Observations	72	72
Log likelihood	-684.090	-46,197.167
Paper-twin FE	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table 8- Impact of Industry Sponsorship on Main Author's Research Output

	POISSON MODEL	
	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-EIGHTED FOLLOW-ON PAPERS
	(8-1)	(8-2)
Industry grant (0/1)	-0.058 (0.081)	-0.091 (0.102)
Number of authors	0.002 (0.004)	0.004** (0.002)
Log(main author's past paper count+1)	0.197*** (0.038)	0.270*** (0.038)
Log(main author's past patent count+1)	-0.077** (0.036)	0.009 (0.043)
Log(main author's academic experience+3)	-0.246*** (0.082)	-0.282*** (0.092)
Main author affiliated with a U.S. institution	-0.075 (0.094)	0.246 (0.152)
Share of authors affiliated with U.S. institutions	-0.148 (0.110)	-0.213 (0.173)
Share of female authors	0.102 (0.101)	0.159 (0.137)
Twin fixed effects	Yes	Yes
Observations	1,236	1,236
Log likelihood	-4,993.648	-36,9643.140
Paper-twin FE	578	578

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1- The Comparison of the Composition of Authorship Team within Twins

	POISSON MODEL			
	DV = NUMBER OF FACULTY	DV = NUMBER OF RESEARCH STAFF (GRAD STUDENTS & POSTDOCS)	DV = NUMBER OF AUTHORS W/ UNIDENTIFIED STATUS	DV = NUMBER OF TOTAL PRE-DISCOVERY PUBLICATIONS OF THE AUTHORSHIP TEAM
	(A1-1)	(A1-2)	(A1-3)	(A1-4)
Academia–industry collaboration (0/1)	0.249 (0.170)	0.055 (0.136)	0.096 (0.127)	0.090 (0.201)
Twin fixed effects	Yes	Yes	Yes	Yes
Observations	72	72	72	72
Log-likelihood	-103.990	-82.281	-147.784	-5,299.343

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table A2- Robustness tests with Negative Binomial and OLS Regressions

	NEGATIVE BINOMIAL		OLS	
	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S LOGGED CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S LOGGED CITATION-WEIGHTED FOLLOW-ON PAPERS
	(A2-1)	(A2-2)	(A2-3)	(A2-4)
Academia–industry collaboration (0/1)	-0.508	-0.427	-0.517	-0.384
	(0.369)	(0.297)	(0.480)	(0.343)
Academia–industry collaboration × Discovery has high commercial potential	0.870**		0.934*	
	(0.391)		(0.5343)	
Academia-industry collaboration × Industry partner is established		1.122***		1.133**
		(0.388)		(0.441)
Number of authors	0.015	0.014	0.014	0.014
	(0.014)	(0.013)	(0.019)	(0.073)
Log(main author's past paper count+1)	0.349***	0.321***	0.340**	0.320**
	(0.112)	(0.089)	(0.154)	(0.152)
Log(main author's past patent count+1)	-0.066	-0.207	-0.103	-0.219
	(0.175)	(0.153)	(0.233)	(0.201)
Log(main author's academic experience+3)	0.022	0.310	0.133	0.355
	(0.461)	(0.342)	(0.575)	(0.449)
Main author affiliated with a U.S. institution	1.309**	1.189***	1.396*	1.180*
	(0.564)	(0.413)	(0.762)	(0.602)
Share of authors affiliated with U.S. institutions	-1.509**	-1.445***	-1.712**	-1.468*
	(0.590)	(0.456)	(0.823)	(0.658)
Share of female authors	-1.666**	-2.522***	-1.688	-2.454**
	(0.770)	(0.852)	(1.057)	(1.060)
Industry grant	-0.411	-0.481	-0.548	-0.522
	(0.483)	(0.470)	(0.543)	(0.518)
Twin fixed effects	Yes	Yes	Yes	Yes
Observations	72	72	72	72
Log likelihood	-581.368	-578.205		
R-Squared			0.763	0.784
Paper-twin FE	33	33	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1

Table A3- Robustness Checks with the Limited Control Set

	POISSON MODEL					
	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S FOLLOW-ON PAPERS	DV = MAIN AUTHOR'S CITATION-WEIGHTED FOLLOW-ON PAPERS
	(A4-1)	(A4-2)	(A4-3)	(A4-4)	(A4-5)	(A4-6)
Academia-industry collaboration (0/1)	0.226*	0.267**	-0.357	-0.243	-0.081	-0.086
	(0.118)	(0.133)	(0.277)	(0.305)	(0.161)	(0.198)
Academia-industry collaboration × Discovery has high commercial potential			0.736**	0.584*		
			(0.300)	(0.335)		
Academia-industry collaboration × Industry partner is established					0.485**	0.502**
					(0.221)	(0.249)
Log(main author's past paper count+1)	0.233***	0.519***	0.205***	0.506***	0.238***	0.494***
	(0.083)	(0.083)	(0.069)	(0.106)	(0.052)	(0.097)
Main author affiliated with a U.S. institution	-0.409**	-0.330*	-0.402**	-0.332	-0.443**	-0.385**
	(0.083)	(0.083)	(0.161)	(0.204)	(0.175)	(0.189)
Twin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	72	72	72	72	72	72
Log likelihood	-282.218	-20,592.058	-267.057	-19,753.719	-273.451	-19,553.834
Paper-twin FE	33	33	33	33	33	33

Robust standard errors are reported in parentheses. The level of analysis is at the twin-paper level.
*** p<0.01, ** p<0.05, * p<0.1