How Does Industry Shape Academic Science? Evidence from "Million Dollar Plants"

Hongyuan Xia

June 25, 2024

Abstract

Firms rely on academic science and actively participate in the production of scientific knowledge. However, the impact of industry on academic science remains unclear. This study utilizes the site selection decisions of "Million Dollar Plants" (MDPs) to estimate the causal effects of industry on academic science. We compare the responses of scientists in counties that successfully attracted MDPs ("winners") with those in counties that narrowly missed out on these MDPs ("runners-up"). We find that MDPs significantly shift scientific research direction without decreasing the quantity or quality of nearby scientists' output. This shift only occurs when MDPs actively fund research or involve a high-skilled labor force. Further investigation indicates that industry not only shapes academic science through direct ties, such as funding or collaboration, but also changes scientists' attention toward more applied and firm-relevant research. Our findings contribute to the understanding of how industry shapes academic science and how scientists react to industry opportunities.

Keywords: University-Industry Relationship, Research Direction, Funding, Attention.

Xia: Cornell Department of Economics (hx276@cornell.edu). This research was supported by a Cornell Center for Social Sciences Grant.

1 Introduction

Firms build on academic science to advance their research and development (R&D) (Cockburn and Henderson 1998; Fleming and Sorenson 2004; Marx and Fuegi 2020; Rosenberg and Nelson 1994), and they are involved in academic scientific knowledge production through collaboration, funding, and technology transfer (Rosenberg et al. 2024, e.g.). However, little is known about the implications of industry on scientific knowledge production, and large-scale causal evidence is scarce. On the one hand, industry can commercialize existing research and provide resources essential for developing new ideas in public science (Babina et al. 2023; Jensen, Thursby, and Thursby 2010; Stephan 2012). On the other hand, firms' emphasis on commercialization can limit academic scientists' contributions to open science and distract scientists from conducting basic research (Czarnitzki, Grimpe, and Toole 2015; Murray 2010; Shibayama, Walsh, and Baba 2012; Toole and Czarnitzki 2010). Academic scientists are typically assumed to have a strong "taste for science" and are often willing to forgo pecuniary benefits in order to engage in open science and select their own preferred projects (Dasgupta and David 1994; Merton 1973; Stern 2004). The potential benefits provided by industry and the distinct norms of academia raise a pivotal question about the interaction between industry and academic science: Can industry shape academic science? If yes, how?

Understanding the impact of industry on academic science is of substantial practical importance. First, local governments in the US have competed fiercely to attract large, high-skilled firms in recent decades by offering tax incentives and other benefits, hoping these firms will create jobs and stimulate economic growth. A recent example of this trend is the competition to attract Amazon HQ2. There are many doubts about whether tax incentives are cost-effective[1]. Universities have been a crucial element in this debate, as Amazon seeks proximity to talent, and university policymakers hope that Amazon's presence will make universities more tech-focused [2]. Empirically examining the impact of industry on academic science can provide evidence for this discussion: Can the impact of these large firms extend beyond the labor market to another domain, academic science, which is considered the engine of long-term growth? Second, universities are increasingly including economic development and commercial translation in their core objectives (Sanberg et al. 2014). However, there is much concern that industry may draw and distract scientists from academia, causing brain

1. Casselman, Ben. "A $2 billion question: Did New York and Virginia overpay for Amazon?", New York Times, Nov 13 2018, https://www.nytimes.com/2018/11/13/business/economy/amazon-hq2-va-long-island-city-incentives.html.

2. Svrluga, Susan. "For universities in Virginia, Amazon's HQ2 came at the perfect moment", Washington Post, Nov 18 2018, https://www.washingtonpost.com/education/2018/11/18/universities-virginia-amazons-hq-came-perfect-moment.

drain and deterring long-term growth. This issue has become more pronounced as tech firms intensify their recruitment of AI experts from universities [3]. Does industry undermine or enhance academic science? Answering this question can help policymakers and university leaders find a balance between academic research and industry involvement.

Previous studies have investigated the relationship between industry engagement and scientists' research output. This literature has explored scientists' commercialization activities, such as patenting and entrepreneurship, suggesting that these activities can alter scientists' research trajectories, but may not decrease academic productivity (Agrawal and Henderson 2002; Azoulay, Ding, and Stuart 2009; Fabrizio and Di Minin 2008; Fini, Perkmann, and Ross 2022; Murray and Stern 2007; Roche 2023). However, these studies focus on scientists who have self-selected to engage with industry (i.e., inventors and entrepreneurs), raising concerns about selection bias. More recent studies have extended this discussion to how industry itself can affect academic science through collaboration and funding (Bikard, Vakili, and Teodoridis 2019; Sohn 2021). These studies predominantly concentrate on a small group of scientists within certain fields who have direct ties, such as funding or collaboration, with industry, leaving a gap regarding how industry shapes the academic science of a large group of scientists across different fields without direct ties with industry. In other words, we know little about how industry shapes the academic research of scientists who are neither industry collaborators nor grantees.

We theorize that industry can shape academic science by shifting academic scientists' research direction toward more applied and firm-relevant subjects. This shift is driven by firm funding and collaboration opportunities. Through resource provision and collaboration opportunities, industry not only directly shapes academic research of scientists with direct ties (e.g., grantees and collaborators), but also indirectly changes scientists' attention toward more applied and firm-relevant research. The refocusing of scientists' attention toward private, for-profit entities is due to the anticipation of potential future resources and collaboration opportunities.

Testing our hypothesis empirically presents multiple challenges. First, industry exposure is not evenly distributed among scientists. Instead, there is significant heterogeneity in its distribution. Scientists located in different areas, working with varied teams, or regarded as "stars" are disproportionately more likely to be exposed to industry (Bikard and Marx 2020; Hsu and Kuhn 2023; Marx and Hsu 2022; Stuart and Ding 2006). This variance makes it

3. Metz, Cade. "When the AI professor leaves, students suffer, study says", New York Times, Sep 06 2019, https://www.nytimes.com/2019/09/06/technology/when-the-ai-professor-leaves-students-suffer-study-says.html.

difficult to isolate scientists' exposure to industry from other influencing factors, such as geographic location and individual prestige. Second, industry may influence outcomes through direct ties (e.g., funding and collaboration) or indirect attention shifts. These changes may happen simultaneously. In short, it is difficult to separate changes in attention from direct access to industry resources and collaboration opportunities.

In this paper, we address the challenges mentioned above by leveraging a quasi-experimental design, the entry of "Million Dollar Plants" into local counties. The MDP entries in our sample come from Conway Analytics. Conway produces the *Site Selection Magazine*, which includes a regular feature titled "Million Dollar Plants" that reports the county that the plant ultimately chose (i.e., the "winner"), as well as the several runner-up counties (i.e., the "losers"). The site selection data has been used by Greenstone, Hornbeck, and Moretti 2010, who study the effect of agglomeration spillovers by looking at the productivity of winners and runner-up/loser counties for Million Dollar Plants (MDPs), and Bloom et al. 2019, who investigate the spillover effects of management practice.

There are several relevant aspects of our research design. First, we rely on the reported location rankings of profit-maximizing MDPs to identify a valid counterfactual for what would have happened to scientists' research production in winner counties in the absence of these MDPs (Bloom et al. 2019; Greenstone, Hornbeck, and Moretti 2010). Since firms typically begin by evaluating dozens of possible locations when they are considering where to open an MDP, the losers in the data are counties that have survived a long selection process but narrowly lost the competition. The identification assumption in our study is that university scientists in the loser counties form a valid counterfactual for university scientists in the winner counties, after controlling individual fixed effects, preexisting trends, and other control variables. Second, several features of MDP entry events enable us to differentiate between the attention mechanism and direct ties mechanisms, such as funding and collaboration. One key feature is that our data include both new MDP entries and existing MDP expansions. If scientists do not actively pay attention to industry, then the impact of new entries and expansions should be similar, as both scenarios provide resources and collaboration opportunities and strengthen the direct ties. Another important feature is our use of the announcement date of MDP entries, rather than the actual entry date. Typically, there is an average two-year gap between these events. Therefore, if scientists actively pay attention to industry, we should see effects immediately following the announcement, rather than after the actual entry of the MDP.

We estimate the causal impact of the entry of "Million Dollar Plants" on nearby university scientists' research output and direction using a difference-in-differences (DID) approach. The data we use (i.e., OpenAlex) covers the full publications of scientists at academic institutions in winner or loser counties between 1990 and 2020. We complement the data by matching patent citations with publications (Marx and Fuegi 2020, 2022). We also link the funding information for each research paper using data from Web of Science. In addition, we obtain university-level data, such as the number of doctoral students and the annual research funding from national institutes or corporations, from IPEDS. We restrict the sample to scientists who were in treated or control counties before the entry of MDPs to avoid potential selection bias caused by the movement of scientists. We also match scientists from the winner counties to those from the loser counties using coarsened exact matching (CEM) to strengthen our causal inference. As a quasi-natural experiment, we focus on the differential effects of MDPs on academic scientists in the winner counties (i.e., treatment) and those in loser counties (i.e., control).

We observe that the entry of "Million Dollar Plants" redirects the research of nearby scientists toward more applied and MDP-relevant fields. Specifically, MDP entries result in a significant boost in commercialization (measured by the number of commercialized papers) and an increase in MDP-relevant research in academic output, without reducing the quantity (measured by the number of published papers) or quality (measured by five-year paper citations) of scientists' work. Following MDP entries, scientists in the treatment group generate an average of 19.7 percent more MDP-relevant papers, and 12.7 percent more papers that are finally commercialized. Additionally, scientists at institutions in winner counties experience an average increase of 0.59 paper citations and 0.02 patent citations, corresponding to 5 and 19 percent increases, respectively. The impact decreases as the geographical distance increases and follows an inverted U relationship with academic age.

In our further analysis of potential mechanisms, we show that the effects are unique to the entry of MDPs that actively fund research and involve a high-skilled labor force. Furthermore, we document a shift in MDP funding toward nearby scientists post entry. We also notice that scientists in the treatment group who had never worked with MDPs before increased their collaborations with MDPs, while scientists who had previously worked with MDPs decreased their collaborations with them. Our analysis supports that industry can shape academic science through building direct ties with scientists. However, the change in research direction persists even after excluding scientists who are MDP collaborators or grantees, suggesting that the main driver may be the redirection of scientists' attention. We provide more direct evidence for the attention mechanism. We find that these shifts occur immediately following the announcement, when the direct ties due to resource provision and collaboration are minimal. Besides, the effects are observed only with the entry of new

MDPs, not with the expansion of existing MDPs. This finding particularly supports our attention mechanism while disputing the direct ties mechanism. If direct ties with industry (i.e., receiving funding from or collaborating with industry) were the main mechanism, the expansion of MDPs should also increase the pool of collaborators and other resources, displaying a strong positive effect. Moreover, the effects are most significant among scientists new to MDP-relevant topics and those without prior commercialization experience, who are most likely to be attracted by the MDPs. Together, these results imply that industry can shape academic science not only through direct ties of resource provision and collaboration, but also by redirecting scientists' attention toward more applied and firm-relevant research. This redirection is based on scientists' expectations of obtaining resources and collaboration opportunities from industry in the future.

This paper makes three main contributions. First, it contributes to the small but growing literature on the impact of industry on academia by providing large-scale causal evidence from a quasi-experiment (Azoulay, Ding, and Stuart 2009; Bikard, Vakili, and Teodoridis 2019; Fini, Perkmann, and Ross 2022; Murray and Stern 2007; Roche 2023; Sohn 2021). The phenomenon of academia-to-industry brain drain, particularly pronounced in the IT sector, has prompted extensive debate (Gofman and Jin 2022). In our analysis, industry positively affects the academic output of scientists, underscoring a complementary relationship between knowledge creation and industry exposure. To the best of our knowledge, this paper is the first to examine the causal impact of industry on academic scientists across different fields.

Second, our study extends the evolving discussion on how scientists select their research topics and direction. Although traditionally, scientists are assumed to choose their research direction freely, forgoing pecuniary benefits (Agarwal and Ohyama 2013; Stern 2004), recent studies have shown that their choices are influenced by personal experience and local contexts (Fini, Perkmann, and Ross 2022; Fry 2023; Koning, Samila, and Ferguson 2021; Sohn 2021; Truffa and Wong 2022). Our work posits that scientists are not disinterested in industry opportunities, rather, they are open to incorporating applied and firm-relevant fields into their research agendas without necessarily compromising research productivity.

Lastly, the geographic nature of our research design enables us to contribute to the understanding of firms' impact on the local economy. Local governments in the U.S. have been competing fiercely to attract large high-skilled firms in recent decades, hoping that these firms will create jobs and generate economic growth. There is much doubt about whether these efforts are cost-effective. Some work has explored the consequences of the entry of large firms on local firms' productivity, management practices, as well as the local labor market (Bloom et al. 2019; Greenstone, Hornbeck, and Moretti 2010; Qian and Tan 2021).

Our study provides a new perspective from university knowledge production to evaluate the implications of large firms on the local economy.

2 Theoretical Background and Hypotheses

2.1 The Interplay between Universities and Industry

The university-industry relationship is bilateral and mutually evolving (Etzkowitz, Webster, and Healey 1998; Rosenberg 1982; Stokes 2011). Recent scholarship increasingly highlights the significant economic contributions of academic scientific breakthroughs and the ways in which firms capitalize on these advancements (Cohen, Nelson, and Walsh 2002). University research improves corporate innovation by cultivating a pool of highly skilled potential employees (Gofman and Jin 2022; Roach and Sauermann 2010; Singh and Agrawal 2011), fostering collaboration between academic scientists and corporate entities (Bikard, Vakili, and Teodoridis 2019; Hausman 2022), commercializing intellectual creations born in academic settings (Agrawal and Henderson 2002), and nurturing startups founded by scholars in academia (Stuart and Ding 2006). In addition, there has been a notable shift in corporate strategies, with companies reducing their investment in fundamental research and increasingly relying on academic innovations (Arora, Belenzon, and Patacconi 2018; Fleming et al. 2019).

While substantial research has been conducted to analyze the effects of academic research on the corporate sector, the influence of the corporate sector on the production of academic knowledge remains less explored and somewhat controversial. Despite academia has being largely an autonomous social ecosystem with its own unique operational methods, principles, cultural values, and resource circulation (Sauermann and Stephan 2013), it remains intricately linked with industry in numerous ways.

Scholars are increasingly investigating the diverse potential interconnections between industry and universities and assessing their influences on the process of academic knowledge production. Studies have scrutinized the influence of industry engagement, such as patenting and entrepreneurship, on academic research output, revealing a spectrum of results (Azoulay, Ding, and Stuart 2009; Fini, Perkmann, and Ross 2022; Merton 1973; Murray and Stern 2007; Sauermann and Stephan 2013). There are instances where the negative consequences of industry engagement are highlighted, suggesting that industrial norms of confidentiality and intellectual property rights might conflict with the academic ethos of widespread publication and dissemination of research findings (Babina et al. 2023; Merton 1973; Stephan 1996).

More recent investigations into academic entrepreneurship have also raised concerns about the "brain drain" of scholars from academia, alongside a decline in local entrepreneurial activities and deteriorating outcomes for PhD students (Gofman and Jin 2022; Roche 2023). On the other hand, earlier studies recognized the positive impact of academic scientists' patenting activities on their research output (Azoulay, Ding, and Stuart 2009; Owen-Smith 2003; Stephan et al. 2007). Some researchers argue that industry engagement may foster remote search and have a positive effect on academic scientists' scholarly production (Fini, Perkmann, and Ross 2022). These studies predominantly concentrate on scientists who have self-selected to engage with industry (i.e., inventors and entrepreneurs), raising concerns about selection bias.

A small but growing literature has extended this discussion to how industry itself can affect academic science through collaboration and funding (Bikard, Vakili, and Teodoridis 2019; Sohn 2021). Bikard, Vakili, and Teodoridis 2019 demonstrates that industry can sometimes foster specialization and boost academic contribution to open science through collaboration with university scientists. A more relevant study, Sohn 2021, finds that universities within a 50-mile radius of the incumbent R&D headquarters show an approximate 28.5 percent increase in industry-relevant research output after the company's entry into agricultural biotechnology R&D. Both studies examine a limited sample of academic scientists or institutions with direct ties with industry and may not provide causal evidence: Bikard, Vakili, and Teodoridis 2019 investigates fewer than 100 scientists who are industry collaborators, and Sohn 2021 only includes academic research in the agricultural biotechnology field and cannot fully eliminate unobserved differences between nearby universities and matched ones. Therefore, there is a gap in understanding whether these results can be generalized to a broader group of scientists, especially those who are not in the agricultural biotechnology field and do not form direct ties with industry. In other words, we know little about how industry shapes the academic research of scientists in different fields and who are neither industry collaborators nor grantees.

2.2 Industry and Academic Research

The previous section has illustrated the knowledge gap in understanding how industry can affect academic science. In this section, we summarize the potential ways that industry can involve in academic knowledge production. Over the past few decades, industry has been involved in academic scientific knowledge production through collaboration, funding, and technology transfer (Rosenberg et al. 2024, e.g.).

Industry can provide funding to academia. The necessity for industry funding has grown as modern science becomes more labor- and capital-intensive and as competition for federal funding intensifies (Jensen, Thursby, and Thursby 2010; Stephan 2012). This type of funding equips scientists with the means to hire personnel and acquire capital goods essential for their research. Industry funding can originate from the market for ideas, where academic outputs are traded for financial benefits (Agrawal 2006; Hegde and Tumlinson 2014; Mowery and Ziedonis 2015; Zucker, Darby, and Armstrong 1998). Alternatively, it may come in the form of long-term investments like gifts, donations, fellowships, and endowed chairs, which are less tied to direct market transactions or contracts.

Industry may actively collaborate with academia and directly employ academic scientists. Such collaboration is frequently promoted nationally, with numerous studies highlighting its strategic value for firms aiming to acquire fresh insights, establish new connections, and enhance R&D productivity (Kafouros et al. 2015; Kaiser et al. 2018; Mindruta 2013; Polidoro Jr, Lampert, and Kim 2022; Tzabbar 2009). In certain scenarios, collaboration may lead to productive specialization between different institutional settings. Instead of infusing academic environments with industry practices, these partnerships may uphold the traditional values of academia by increasing publications and reducing patenting rates (Bikard, Vakili, and Teodoridis 2019). Industry also offers other forms of support to academia. The literature has shown that informal interactions with industry professionals can promote the diffusion of knowledge (Andrews 2019; Hasan and Koning 2019; Roche, Oettl, and Catalini 2023).

The resources and collaboration opportunities provided by industry are frequently bounded by geographical constraints. A number of studies have found that geographical proximity is associated with a greater likelihood of market transactions between industry and university (Agrawal 2006; Zucker, Darby, and Armstrong 1998). Geographical proximity may facilitate the spillover of industry information and knowledge to academic scientists, to the extent to which the informal and labor market networks of scientists are embedded in the local community (Atkin, Chen, and Popov 2022; Catalini, Fons-Rosen, and Gaulé 2020; Roche, Oettl, and Catalini 2023).

2.3 Scientists' Reaction: Resources, Collaboration and Attention

The geographical proximity to industry enhances the visibility and feasibility of resources and collaboration opportunities from industry, yet how does industry shape academic science remains unclear. In this section, we develop theory on how industry can shape academic scientists' research trajectories.

On the one hand, prior research suggests that academic institutions possess a relative competitive advantage in investigating more basic, fundamental questions, whereas firms maintain a relative advantage in development and commercialization (Agarwal and Ohyama 2013; Roach and Sauermann 2010). The comparative advantage is due to resources and expertise in fostering efficient commercialization (Bush 1945; Merton 1973; Rosenberg and Nelson 1994; Sauermann and Stephan 2013). Aligned with this assumption, academic scientists are typically assumed to have a strong "taste for science" and are often willing to forgo pecuniary benefits in order to engage in open science and select their own preferred projects (Dasgupta and David 1994; Merton 1973; Stern 2004). Moreover, scientists often find themselves constrained within the confines of their academic disciplines (Kuhn 1962), a phenomenon further emphasized by the burden of knowledge theory (Jones 2009). Furthermore, the academic institutional environment does not consistently reward extensive, exploratory searches (Chai 2017). Peer review processes and grant evaluations tend to favor adherence to a limited, established set of scientific problems, a trend reinforced by journal rankings and influential scientists who set agendas within disciplines (Azoulay, Fons-Rosen, and Zivin 2019). Recent research also indicates that scientists engaging in broader, more exploratory searches often face career and promotional disadvantages (Hofstra et al. 2020). Therefore, industry may not be able to shape academic science as scientists may not react to the opportunities enabled by industry, considering that pursuing these opportunities often requires distant search in firm-relevant domains, which may not be rewarded by academia and conflicts with their academic interests.

On the other hand, academic scientists' choices of research topics are influenced by nearby events and organizations (Fry 2023; Sohn 2021; Truffa and Wong 2022). As mentioned before, industry can provide resources (e.g., funding) to scientists and may actively interact with scientists in academic collaboration. Scientists may change their research direction because the funding they receive requires them to solve applied questions. They may also change direction through exposure to firm-relevant ideas via collaboration or informal interactions with industry labor force. These changes typically occur after scientists have received funding from or interacted with industry, both formally and informally. Can scientists react to industry even before they receive funding or collaborate with industry labor force, or without these direct ties? In this paper, we propose that industry can influence the salience of firm-relevant topics, thereby increasing the attention devoted to these problems.

The concept of attention or "focus" emerges as a crucial endogenous mechanism within the scientific production function and is shaped by both internal academic motivations and external socio-economic and environmental factors (Rosenberg and Nelson 1994; Rosenberg et al. 2024). The resources and collaboration opportunities from industry are important as modern science is becoming more labor- and capital-intensive and as competition for federal funding intensifies (Jensen, Thursby, and Thursby 2010; Stephan 2012). The proximity to industry increases visibility and feasibility of these opportunities, prompting scientists to consider the potential benefits of aligning their research with industry interests. Therefore, scientists may intentionally pay more attention to and work on firm-relevant topics in anticipation of obtaining resources and collaboration opportunities from industry in the future.

For instance, the establishment of Amazon's second headquarters in Crystal City presents opportunities for scientists at Virginia Tech and George Mason University. Amazon has gathered talent in computer science and data science, maintained high-performance computing resources, and actively funded research over the past decade. For academic scientists, collaborating with Amazon or receiving its funding or resources can not only foster novel findings but also broaden their impact beyond academia. However, to collaborate or interact with Amazon and obtain its resources, scientists need to focus more on topics relevant to Amazon. A scientist working on supply chain optimization is more likely to obtain resources from and interact with Amazon in the future than a scientist working on analytical modeling.

Moreover, the redirection of research toward more applied fields does not inherently diminish its academic value. First, as suggested by the theory of remote search and exploration (March 1991), diversity in knowledge and the formation of unconventional combinations often lead to significant innovations (Uzzi et al. 2013). Second, although new research trajectories in public science may not align perfectly with a scientist's initial academic focus (Azoulay, Ding, and Stuart 2009; Evans 2010), commercial engagements can offer valuable learning opportunities, especially when they align with or complement existing research projects (Stephan et al. 2007).

Therefore, we propose the following two hypotheses:

Hypothesis 1. The geographical proximity to industry influences nearby scientists' research trajectories, shifting research direction toward more applied and firm-relevant fields.

Hypothesis 2. The geographical proximity to industry does not decrease nearby scientists' research productivity, instead, they can improve the research quality.

As we discussed above, industry can shape academic science not only through direct ties of resource provision and collaboration but also through indirect attention shifts in anticipation of future resources and collaboration opportunities. In both direct and indirect mechanisms, the availability of resources and collaboration opportunities plays a crucial role in redirecting scientists to firm-relevant fields.

Hypothesis 3. The geographical proximity to industry shifts scientists' direction toward firm-relevant fields only when industry offers resources and collaboration opportunities.

Furthermore, to distinguish scientists' own attention shifts from direct ties of resource provision and collaboration, we implement several tests based on our research design, the entry of "Million Dollar Plants". First, scientists are likely to be more attentive to new industrial firms than to the expansion of existing ones, despite both increasing the pool of collaborators and other resources. Second, the announcement time of firm entry events is different from the actual entry time. We expect fewer collaborations and less resource provision between firms and academic scientists during the gap between the announcement and the actual entry.

An academic scientist's previous connection with industry might moderate the impact of industry. While industry can affect nearby scientists through direct ties of resource provision and collaboration as well as indirect attention shifts, different mechanisms may work on different scientists in different ways. If direct ties of resource provision and collaboration are the main drivers for the change in scientists' research direction, we may expect scientists with prior experience in firm-relevant fields or commercialization to be more likely to change their research direction, as firms are more likely to allocate resources to those who have solved practical industry questions. Conversely, if the indirect attention shift is the main mechanism, we expect scientists without prior experience to be more likely to change their research direction, as those with prior experience have already paid attention to these topics. For instance, a scientist already working with a pharmaceutical company may continue focusing on drug development research and remain unaffected by new industry opportunities in this field.

Hypothesis 4. The geographical proximity to industry influences nearby scientists' research trajectories, shifting research direction toward more applied and firm-relevant fields primarily in response to new firm entries. This shift occurs before the physical establishment of the firm and is moderated by the scientists' previous connection with industry.

3 Data and Research Design

In estimating the effects of industry on academic scientists, researchers have identification challenges. Ideally, we would like to randomly place firms to a random set of academic scientists, or randomly put academic scientists in regions with and without firms, and observe follow-on publication and commercialization efforts of academic scientists. However, firms are always not randomly scattered, in reality, both firms and scientists can choose their location, raising the concern of selection bias. First, firms often select physical locations that will enable them to exploit local university spillovers and the scientists in these regions may be more likely to pursue industry opportunities and investigate topics that are of interest to firms. Second, scientists with a higher interest in commercialization may select regions with the presence of R&D-intensive firms. Besides, the discoveries from academic labs can serve as input for firm innovation through knowledge diffusion and attract firms to nearby regions, causing reverse causality.

To deal with these identification challenges, this paper uses reported location rankings of profit-maximizing "Million Dollar Plants" (MDPs, hereafter) to identify a valid counterfactual for what would have happened to university scientists' research production in winner counties in the absence of MDPs. This approach is inspired by Greenstone, Hornbeck, and Moretti 2010, who study the effect of agglomeration spillovers by looking at the productivity of winners and runner-up/loser counties for MDPs. One recent example is Amazon: Over 200 counties competed for the opportunity to host Amazon's second headquarters, offering various tax incentives and other benefits. The competition narrowed down to a shortlist of 20 counties, with Amazon representatives visiting these locations. In the final stages, a few contenders, notably Crystal City, Long Island City, and Dallas, engaged in detailed discussions with Amazon regarding the HQ2 location. Up until October, Dallas appeared to be a strong candidate, however, Crystal City and Long Island City ultimately secured the position for Amazon's second headquarters.

This scenario with Amazon illustrates that researchers can establish an effective control group using the final standings in corporate decision-making processes, despite not having complete insight into all the factors, which are often unobservable, influencing such decisions. It's true that many observable and unobservable factors are taken into account in Amazon's decision process, and some of them, such as the talent of local universities, are related to the academic scientists' research. However, the winners and losers are chosen from an initial sample of more than 200 counties. The losers are counties that have survived a long selection process but narrowly lost the competition. Even if this assumption fails to hold, we presume that this pairwise approach is more reliable than using regression adjustment to compare the research trajectories of academic scientists in counties with MDPs to the other 3,000 U.S. counties or using a matching procedure based on observable variables.

The use of winner and loser counties helps us control the strategic selection into particular regions by MDPs and academic scientists, and hence isolates the effect of MDPs. In addition, we focus on academic scientists who have been in the counties before MDPs and stayed there after the entry of MDPs to avoid selection bias from academic scientists. We assume that academic scientists in winner and loser counties are similar or follow a parallel trend in terms of academic knowledge production and commercialization. The subsequent analysis provides evidence that supports the validity of this assumption.

3.1 Data

We implement the research design using data from Conway Analytics. Conway is a company that offers services to governments and companies, helping them decide where to locate new establishments. Conway produces the *Site Selection Magazine*, used in Greenstone, Hornbeck, and Moretti 2010 and Bloom et al. 2019. We obtain 105 pair records where a firm reports the county that the MDPs chose (i.e., the "winner") and the runner-up county or counties (i.e., the "losers") from 2000 to 2015. We extend the original MDP data and manually collect announcement date, opening date, latitude and longitude, and purpose by searching on Google using the name, county, and announcement year of each MDP.

To understand how the entry of MDPs affects local academic scientists' output, we compile a large dataset using rich publication information from various sources. The core information on the sample underlying this study is OpenAlex publication data (Priem, Piwowar, and Orr 2022), which provides comprehensive information about authors, citations, institutions, published journals, and concepts. We obtain the publication information of academic scientists who work in academic institutions located in winner or loser counties. We drop those who leave the institution before the entry of MDPs and those who come to the institution after the entry of MDPs to avoid potential selection bias from academic scientists. We only keep the information of publications that are not 10 years earlier or later than the entry of MDPs to enhance the comparability between pre- and post-event samples.

Once we obtain the publication information of academic scientists in both winner counties and loser counties, we supplement the data by linking patent citations and patent-paper pairs information to each publication using the "Reliance on Science" data by Marx and Fuegi 2020, 2022. Patent citations to papers indicate that the inventors incorporated the knowledge presented in these papers into their process of invention (Roach and Cohen 2013). Patent-paper pairs characterize whether the paper is eventually commercialized in a patent, which is a common tool to measure commercialization (Koffi and Marx 2023, e.g.). We also collect patent data from USPTO to determine whether a paper is MDP-relevant. The next section explains in detail how we identify MDP-relevant publications using the patent data and academic scientists' publication information. Finally, we obtain funding information for each paper from Web of Science (WOS), which enables us to track whether a scientist received funding from MDPs.

We take a Coarsened Exact Matching (CEM) approach to match each treated academic scientist with a control scientist. CEM matches samples based on ex-ante criteria to ensure that one does not confound treatment effects with observable heterogeneous pre-trends, while effectively minimizing heterogeneity between the observations in the treatment and control groups and therefore strengthening causal inferences. The treatment and control scientists were one-to-one matched within the same MDP pair (i.e., winner and loser counties of the same MDP entry event) according to the following criteria: (1). the number of MDP-relevant publications, (2) the number of patent-cited publications, (3) the mean of paper citations, and (4) the mean of patent citations. Overall, our effort results in an unbalanced panel dataset of 1,817,585 scientist-year observations.

3.2 Empirical Strategy

Estimation Equation. We estimate the causal impact of "Million Dollar Plants" on local academic scientists' research output and direction using a difference-in-differences (DID) approach. Specifically, we estimate the following equation:

$$Y_{i,t} = f\,(\alpha + \beta Treatment_i * PostEntry_{it} + Controls_{i,t} + \lambda_p + \eta_i + \gamma_t),$$

Where $Y_{i,t}$ is the measure of outcome, for example, the number of publications of scientist i in year t. $Treatment_i$ is a binary variable that equals one if the scientist is in a university located in winner counties. The winner and loser counties information is from the *Site Selection Magazine* of Conway Analytics. $PostEntry_{it}$ is an indicator variable that switches to one in the year that follows the MDP's decision and announcement to enter the county. For the scientists in the control group, even though there is no actual entry of the MDP, we assign the same value for $PostEntry_{it}$ as those in corresponding winner counties. We include the pair fixed effects λ_p, where a pair is defined as the MDP with both winner and loser counties. The inclusion of λ_p ensures that all the comparisons are between the scientists in the winner and loser counties of the same shock of MDP entry. In other words, we do not rely on comparisons across different MDPs and only compare scientists who are affected by the same MDP. The scientist fixed effects η_i control for many time-invariant characteristics that determine an academic scientist's research productivity and industry connection, such as academic reputation. The year fixed effects γ_t control for the overall time trend. β is the coefficient of interest in this study and captures the treatment effects of MDPs.

$Y_{i,t}$ in the main analysis include several outcome variables, and some of them, such as the number of publications, are discrete, non-negative count variables. For these variables, we use the conditional fixed effects Poisson model with QML (quasi-maximum-likelihood) standard errors clustered at the author level. For continuous outcome variables, for example, the mean of patent citations, we perform an ordinary least squares (OLS) estimation.

Dependent Variables. We use OpenAlex data to construct both quantity and quality of academic scientists' output. The quantity of academic scientists' research is measured by counting the number of publications of scientist i published in year t. The quality of academic scientist i's research is measured by the mean of the five-year forward citations of papers published in year t. Citations serve as a measure of the intellectual value that the academic community assigns to a publication (Leahey, Beckman, and Stanko 2017).

We are interested in the research direction of academic scientists. The variable we use to approximate the research direction of scientists is the number of MDP-relevant publications. We identify MDP-relevant publications using a data-driven approach. For each MDP, we take advantage of its parent firm's public patent records from USPTO and merge the patent with the patent-to-paper citations data that we have mentioned before. We assume that firms cite the papers in their patents if they find these papers relevant and useful in their R&D process (Arora, Belenzon, and Patacconi 2018; Arora, Belenzon, and Sheer 2021). For each cited paper, we assemble its concepts using the concept information provided by OpenAlex. The concepts are abstract ideas that the papers are about, and there are around 20,000 concepts. We gather and rank the layer-2 concepts of all the publications that the parent firm of MDP cite, and we keep the top 20 frequently cited concepts as MDP-relevant concepts. Finally, we define MDP-relevant publications as papers with MDP-relevant concepts, and calculate the number of MDP-relevant publications for each scientist i in year t. This variable measures whether scientists shift their direction toward fields that the MDP is interested in. More detail on the construction of MDP-relevant papers is provided in Appendix A.2.

The last set of dependent variables we construct pertains to the commercialization of academic research. We link patent-to-paper citation data and patent-paper pair data (Marx and Fuegi 2020, 2022) to the publication data obtained from OpenAlex based on the identifier of each publication. Using the linked data, we construct several variables. First, we use the patent-paper pair data to create a variable characterizing how many papers published in year t are commercialized in a patent-paper pair by scientist i. A paper has a patent-paper pair if it is cited by a patent where the author(s) of the paper were also inventors on the patent and the assignee of the patent is a firm. Appendix A.3 provides an example of the patent-paper pairs. This variable quantifies the actual commercialization of academic research. In addition to actual commercialization, we use patent citation data to capture the latent commercialization potential inherent in scientists' work. Patent citations to scientific papers have been used to understand how inventors and scientists search for commercializable basic science (Bikard and Marx 2020; Fleming et al. 2019). The second variable we construct is the number of papers published in year t that are cited by patents within five years. For each publication, we check whether it has been cited by any patent within five years of its publication. We aggregate this to the author level by counting how many papers published in year t are cited by patents. Third, we calculate the mean of the five-year patent citations of papers published in year t.

Moreover, to test the funding mechanism, we collect papers' funding information from Web of Science (i.e., WOS) and construct a variable indicating the number of papers that scientist i received grants from the MDP in year t. In addition, we construct another variable to proxy for collaboration with MDPs and other firms. OpenAlex contains information about coauthors and their affiliations for each publication. Based on the information, the variable we construct is the number of papers published in year t that are coauthored with the MDP.

Independent Variables. The independent variables in our study are $Treatment_i$, a binary variable that equals one if the scientist is in a university located in winner counties, and $PostEntry_{it}$, an indicator variable that switches to one in the year that follows the MDP makes the decision and announces to enter the county.

Moderators. First, we are curious about the heterogeneous effects across different MDPs. We construct a binary variable indicating whether the MDP entered this county with a low-skilled plant. We define low-skilled plants as plants hiring workers doing routine jobs, such as call service. We verify each of the MDP entry events by searching on Google. We also construct an indicator which equals one if the parent firm of the MDP actively funded research. Besides, the Conway data contains information about whether the MDP entered the county for the first time or the MDP just expanded the existing plant.

Second, we investigate how the scientist's prior experience in MDP-relevant topics, commercialization, and industry connections moderate the effects of MDP. We construct four moderators: (1). an indicator which equals one if the scientist has never worked on MDP-relevant concepts before the entry of MDPs; (2). a binary variable which equals one if the scientist has fewer than three publications that are MDP-relevant before the entry of MDP; (3). whether the scientist has commercialized papers before; (4). whether the scientist has coauthored with any firms in fewer than three papers before.

Third, we are interested in how the effects of MDP differ across locations, universities, and scientists. We calculate the linear spatial distance between the academic institutions and the MDPs. OpenAlex provides the latitudes and longitudes of every academic institution, and we supplement it by manually collecting the latitude and longitude of every MDP from Google Map. We calculate the academic age of each scientist, which is defined as the time period between year t and the year of the scientist's first publication. We also construct an indicator which equals one if the scientist is working in a R1 university.

Control Variables. All estimation models add pair fixed effects λ_p, scientist fixed effects η_i, and year fixed effects γ_t. We also include time-variant control variables at the institution level and the scientist level.

At the institution level, the scale and R&D grants may affect the direction and intensity of academic science (Babina et al. 2023). We therefore control for the number of earned doctorates, total R&D grants, and the number of published papers at the university-year level. In addition, as we also examine the commercialization of academic science, the source of R&D grants, universities' prestige in corresponding field and the institutions' prior commercialization experience may confound our estimation (Babina et al. 2023; Fleming et al. 2019). We thus control for the amount of R&D grants from business corporations, NSF, and NIH, as well as the number of patent-cited papers and the number of MDP-relevant papers at the university-year level. We obtain the R&D grant data and number of earned doctorates every year from Integrated Post-secondary Education Data System (IPEDS).

At the scientist level, since we have already added scientist fixed effects, we do not need to add any variables that are invariant across years. Therefore, we only control for the number of papers and the number of different concepts at the scientist-year level.

4 Results

4.1 Descriptive Statistics

We provide summary statistics for our main variables in Table 1. On average, one scientist publishes 3.13 papers in one year, however, only 0.01 papers are commercialized, only 0.03 papers are MDP-relevant, and only 0.10 papers are cited by patents within five years. Besides, we find scientists receive 11.26 paper citations and 0.10 patent citations on average for their papers published in a specific year. Our findings are consistent with the fact that small fraction of the academic work is exploited by firms and less than five percent of all academic publications are cited by patents (Ahmadpoor and Jones 2017).

The list of all MDPs and universities involved in our research is relatively long. Most of the MDPs are established public firms from different industries. Some of the most frequent institutions include Arizona State University, Carnegie Mellon University, Columbia University, Duke University, Georgia Institute of Technology, Indiana University, Iowa State University, Louisiana State University, University of South Carolina, North Carolina State University, Ohio State University, Pennsylvania State University, Southern Methodist University, Stony Brook University, Syracuse University, University of Alabama, University of Central Florida, University of North Carolina, and University of Texas at Austin.

In Table B.1, we show difference of means test according to whether the scientist is in an institution in a winner (i.e., treatment) county or a loser (i.e., control) county before the entry of MDP. Despite the relatively modest magnitude of the differences between treatment and control groups, our analysis, bolstered by Coarsened Exact Matching (CEM), reveals statistically significant distinctions in some variables. To further address the imbalance concern between treatment and control groups, we employ a regression analysis. This approach allows us to compare the treatment and control groups while meticulously controlling for a range of control variables as well as pair, scientist, and year fixed effects, as outlined in our main specification. We find, as displayed in Table 3, that prior to the entry of these MDPs, there are no significant differences among key dependent variables. These results justify our research design by showing there is no significant pre-treatment difference between treatment and control groups after controlling some confounders. Moreover, we display that MDPs' decisions to choose winner counties over loser counties are not related to the number of papers, number of MDP-relevant papers, and the number of commercialized papers before the entry of MDPs at the county level, as shown in Table B.2.

4.2 Main Results

Table 4 presents the two-way fixed effects regression estimates of the impact of "Million Dollar Plants" on academic scientists' research quantity and quality. Columns (1) to (3) display the impact of MDP on academic scientists' research quantity, measured by the number of papers published in year t, and Columns (4) to (6) uses the mean of five-year forward citations to papers published in year t to examine the effects on the quality of scientists' output. Columns (1) and (4) include scientist and year fixed effects to control for time-invariant characteristics of academic scientists and overall time trend. Columns (2) and (5) add pair fixed effects, in addition to scientist and year fixed effects to ensure that all the comparisons are between the scientists in the winner and loser counties of the same shock of MDP entry. Columns (3) and (6) contain pair, scientist, and year fixed effects as well as institution-level control variables to avoid potential confounders from R&D grants and institution prestige. We do not include scientist-level control variables in Column (3) as the dependent variable is the number of papers, which is the same as the number of papers and highly correlated with the number of different concepts that we add in other models. We use the conditional fixed effects Poisson model for columns (1), (2), and (3) since the outcome variable is discrete and non-negative, and we use linear model for the five-year paper citation variable, which is continuous.

Columns (1) to (3) do not show any evidence that the entry of “Million Dollar Plants” increases the quantity of academic scientists’ research significantly, and this finding is robust across different model specifications. Columns (4) to (6) suggests that the entry of MDPs increases the quality of the research. Column (6) shows that five-year paper citations increase 0.60 on average for scientists in the institutions located in winner counties. Considering the average five-year paper citations are 11.26 in our sample as shown in Table 2, our findings suggest MDPs increase the five-year paper citations by 5.3 percent. This result complements existing findings that suggest academic research and commercialization are not substitutes to each other (Fabrizio and Di Minin 2008, e.g.), even though the existing literature focuses on self-selected industry engagement instead of the effects of industry itself.

Table 5 focuses on the impact on the commercialization and direction of academic research. Columns (1) to (3) show the scientists’ shifts in research direction toward MDP-relevant research in response to the entry of “Million Dollar Plants”. The outcome variable is the number of MDP-relevant papers that are categorized by OpenAlex with concepts often cited by the MDP. Columns (4) to (6) use the number of commercialized papers based on the scientists’ research published in a certain year. We use patent-paper pairs to measure whether a paper is finally commercialized, which is a direct measure of scientists’ engagement in commercialization. The number of observations is different across models as the Poisson estimation will drop the scientists that never transfer research to patents or publish MDP-relevant papers.

We find the number of MDP-relevant research papers has increased by 19.7 percent for scientists in universities located in winner counties, relative to scientists in the control group. This finding suggests scientists shift their research direction toward fields that are of interest to firms. We also find academic scientists in the treatment group publish 12.7 percent more commercialized papers on average. This evidence suggests the entry of “Million Dollar Plants” shifts scientists’ direction toward MDP-relevant topics and fosters the commercializable research of scientists and increases the commercialization of academic research. Ahmadpoor and Jones 2017 show that less than five percent of all academic publications being cited by patents. Existing research papers mainly focus on the demand side of commercialization and explore how firms allocate attention to scientific literature (Bikard and Marx 2020; Fleming et al. 2019). This study investigates the supply side of commercialization, the production of applied and commercializable research, and highlights the importance of the local industry. All the results are robust in linear model specification and extensive margin specification as shown in Table B.4 and B.5.

We also graphically examine the validity of the parallel trends assumption in the DID analysis. Figure 1 plots the coefficients from a conditional fixed-effects Poisson regression, which regresses the number of MDP-relevant publications and the number of commercialized papers on a set of interaction terms between treatment indicator and time indicator variables. The graph supports the absence of confounding pre-treatment trends.

We further investigate the impact across different groups. In the rest of this section, we display how the impact of "Million Dollar Plants" varies across distance, academic age, and different types of universities. Column (1) of Table B.3 shows that the differential effects of "Million Dollar Plants" on scientists with different academic ages. Academic age is defined as the time period between year t and the year of the scientist's first publication. We find there is an inverted-U relationship between academic age and the impact of "Million Dollar Plants", which suggests middle-age scientists are more likely to be affected the MDP. This result can be explained by the nature of academic knowledge production and commercialization activities: middle-age scholars have much freedom to choose research topics and have energy to pursue these topics (Azoulay et al. 2020). Our results in Column (2) of Table B.3 suggest that the effects of MDPs on research direction decrease with distance between the MDP and the scientist's affiliated institution. Column (2) shows the effects on the number of MDP-relevant papers decrease even though the coefficient is not statistically significant. We admit that the geographical proximity may include some factors such as informal interactions. Finally, we examine the heterogeneous effects of MDPs across different types of universities. Column (3) of Table B.3 shows that scientists in both R1 and Non-R1 universities increase to publish in MDP-relevant fields.

4.3 Mechanism Tests

In this section, we examine the mechanisms that drive the effects of "Million Dollar Plants" on scientists' research direction toward MDP-relevant topics. As we discussed in section 2, the entry of MDPs can shift scientists' research direction through direct ties of resource provision and collaboration, as well as attention change. However, this attention mechanism is hard to distinguish from the effects driven by firms' direct support to involve scientists, such as funding and collaboration. In the following part of this section, we provide direct and indirect evidence for these mechanisms.

In our first step, we ask whether any MDPs can shift scientists' research direction, or are only certain kinds of MDPs capable of shifting direction? Table 6 examines what kind of MDPs can have effects on scientists' research direction. Columns (1) and (2) implement sub-sample analysis based on whether the MDP involves a high-skilled or low-skilled labor force. The high-skilled and low-skilled status is defined by the purpose of the plant and whether the plant only hires low-skilled workers, with a calling service center serving as an example of a low-skilled plant. We do not find that low-skilled MDPs have effects on scientists' research direction, as displayed in Column (2). Columns (3) and (4) implement another sub-sample analysis based on whether the parent firm of the MDP actively funds research or not. Our results indicate that the entry of MDPs whose parent firms actively fund research increases scientists' MDP-relevant papers by 17.4 percent. These effects are not significant for the entry of MDPs whose parent firms do not actively fund research. These findings suggest that not all MDPs can change scientists' research direction toward MDP-relevant topics, instead, only MDPs that involve a high-skilled labor force and might fund academic research can have an effect on research direction. In other words, in the absence of funding incentives and collaboration opportunities, scientists will not change their research direction in response to the MDP entry.

While Table 6 highlights that funding, collaboration or informal interactions with firm labor forces might be potential explanations for the MDP effects, it is not entirely clear whether the change in research direction is driven by actual funding and collaboration or by a shifted attention toward funding and collaboration opportunities. Table 7 tests the effects of direct ties of resource provision and collaboration. Column (1) of Table 7 shows that treated scientists are more likely to receive grants from MDPs compared to their controls. After the entry of MDPs, scientists in the treated counties produce 21.4 percent more papers funded by the MDPs, although this effect is only significant for a small group of scientists. Column (2) shows that treated scientists are not more likely to engage in collaboration with the MDP compared to their controls. Instead, we find that scientists in the treatment group have fewer coauthored papers with the MDP after the MDP's entry. Column (3) further examines the change in collaboration with MDPs, showing that scientists who have never coauthored with the MDP before produce more MDP-coauthored papers after the MDP's entry, while scientists who had previously coauthored with the MDP collaborate with the MDP less frequently after the MDP entry. Prior research suggests that university–industry collaboration can sometimes foster specialization and boost academic contribution to open science (Bikard, Vakili, and Teodoridis 2019). Based on this theory, scientists who once collaborated with the MDP are more likely to specialize in academic science rather than involving themselves in commercialization activities after the entry of the MDP. Therefore, they are more likely to focus on academic publications and less likely to coauthor with the MDP to explore applied topics. Column (4) of Table 7 supports this prediction.

Despite finding evidence supporting the funding and collaboration mechanism, we also notice that there are very few MDP grantees and collaborators. For example, only 1,737 scientist-year observations in the collaboration estimation, which accounts for 0.1 percent of our sample. Columns (5) to (7) of Table 7 exclude all the scientists who once collaborated with the MDPs, received grants from the MDPs, or both, either before or after the MDP entry. We find that the results still hold, suggesting that direct resource provision (i.e., funding) and collaboration (i.e., coauthorship) may not be the main mechanisms driving our results.

We further test the existence of our attention mechanism. The first test we conduct is based on the feature of our MDP entry data. Instead of using the actual entry year of the MDP, we use the announcement year. On average, there is a gap of more than two years between the actual entry date and the announcement date, based on our manual search, as displayed in Figure 2. Our results in Figure 1 show that the effects start right after the announcement instead of waiting until the actual entry of MDPs. While Figure 1 suggests the effects start right after the announcement, there are several concerns: First, we approximated the time gap using the calendar year, so a paper published at the beginning of a year has the same time stamp as a paper published at the end of a year. Second, while the scientists' attention may shift immediately, the publication takes time ranging from months to years. If the effects appear on the next day of the announcement, it may be due to some pre-existing factors.

To address these concerns, we collect the announcement date, the opening date of MDPs, and the publication date of each paper and calculate the time gap using the exact dates instead of calendar year. Figure 3 provides comparison of the county-level number of MDP-relevant papers and the scientist-level average number of MDP-relevant papers between treatment and control groups. The average number of MDP-relevant papers begins to diverge one year after the announcement and before the average entry date of MDPs. Figure 4 re-estimates the event-study model using the date-calculated and more granular time gaps (i.e. half-years and quarters). We find the effects happen after at least six months but before thirty months of announcement. Lastly, we implement a placebo test by using the opening date instead of the announcement date of MDP to calculate the time gap. We notice that scientists in winner counties produce more MDP-relevant papers before the actual opening date, as displayed in Figure 5. Combined together, these findings suggest that the research direction shift of scientists begins before the MDPs interact formally and informally with the local scientists.

Table 8 displays a sub-sample analysis based on whether the MDP entered the county for the first time or expanded an existing plant. Column (2) suggests that the expansion of MDPs has no significant effect on the change in scientists' research direction (p value $= 0.257$). This finding particularly supports our attention mechanism while disputing the direct ties mechanisms. If direct ties with industry (i.e., receive funding from or collaborate with industry) are the main mechanism, the expansion of MDPs should also increase the pool of collaborators and other resources, displaying a strong positive effect.

Table 9 presents the heterogeneous impact of MDP on scientists with varying degrees of prior commercialization experience. Academic scientists who had never worked on MDP-relevant concepts are significantly more likely to switch to MDP-relevant concepts after the entry of MDPs compared to those who once worked in these fields (p value < 0.01), as suggested in Column (1). Column (2) displays similar results for scientists who rarely worked on MDP-relevant concepts, defined as those with fewer than three MDP-relevant papers, before the entry of MDP. Similar results have also been found in scientists who have never commercialized papers before or who have rarely worked with firms previously. These results suggest that scientists who are novices in the MDP-relevant or applied fields are more likely to switch their direction to these fields, while those with prior experience react much less actively to the entry of MDPs. This finding supports the attention view since scientists with prior experience in MDP-relevant research have already paid attention to these research topics and pursued commercialization opportunities. Therefore, they are less likely to react to the opportunities enabled by MDPs.

4.4 Alternative Explanations and Robustness Tests

Our results are also robust across different specifications. First, we implement our main analysis using a linear model instead of a conditional fixed effects Poisson model, which drops the scientists without any MDP-relevant publications or commercialized publications. The results are shown in Table B.4 and are similar to our main specification. Second, we estimate the extensive margin (e.g., whether the scientist engages in MDP-relevant research) instead of the intensive margin (e.g., the number of MDP-relevant papers) in our main specification. The extensive margin estimation in Table B.5 shows significant and comparable results.

One potential violation of our empirical analysis is the presence of omitted variables that affect both firms' MDP decisions and scientists' research trajectories simultaneously. For example, a breakthrough in mRNA technology may encourage biotech firms to set up MDPs near universities with biomedical expertise. At the same time, biomedical scientists are more likely to pursue mRNA-related topics. In other words, the observed change in research direction of scientists may not be due to the MDP, but rather a result of a common industry shock. To alleviate this concern, we construct a variable characterizing industry-relevant research. Instead of using the concepts most cited by the MDP, we use the concepts most cited by all the firms in the industry of the MDP. We find that the results are not significant, as displayed in column (1) of Table B.6. This serves as a falsification test, indicating that scientists only change research direction toward the MDP, not general industry opportunities. Another possible scenario is that the treated universities intend to transit to a more technology-focused mission, thereby trying to attract MDPs and encouraging scientists to pursue applied research. We address this concern by examining scientists' collaboration with any firms. If the universities' mission of transition is the driving factor, we would expect to see increased collaboration with any firms, not only the MDPs. However, we do not observe such effects. Columns (3) and (4) of Table B.6 use two other measures—the number of patent-cited papers and the mean of patent citations, respectively—to measure the commercialization potential of the papers, and yield similar results.

Furthermore, to ensure our results in Table 9 are driven by prior experience instead of age, we rerun the estimation by adding academic age fixed effects. Some may worry that the results showing scientists without prior commercialization experience are younger and more likely to engage in MDP-relevant research simply because they have more free time compared to senior scientists. In fact, the correlation coefficients between academic age and our four measurements of prior experience range from -0.03 to -0.13, suggesting there is considerable variation in the relationship between academic age and prior experience. Moreover, our robustness test in Table B.7 shows similar results after adding academic age fixed effects.

5 Discussion and Conclusion

In this paper, we investigate how industry shapes academic science. We specifically assess the impact of industry on the research direction and commercial focus within academic scientists' publications. This investigation is critically important in the context of growing concerns regarding the influence of industry on academic standards and practices. Our underlying hypothesis posits that, in response to the geographical proximity to industry, scientists will shift their research direction toward addressing firm-relevant problems. They are expected to explore solutions pertinent to industry, thereby broadening the scope of their research. This shift is anticipated to amplify the significance of their research as they integrate interdisciplinary approaches and concepts, introducing new dimensions to their fields.

We face empirical challenges in investigating the impact of industry on academic scientists because neither firms nor scientists randomly sort into regions that involve interactions between firms and scientists. This paper attempts to overcome the challenge of selection bias by using reported location rankings of profit-maximizing "Million Dollar Plants" to identify a valid counterfactual for what would have happened to university scientists' research production in winner counties in the absence of MDPs. We analyze 105 MDP entry pairs with reported winner and loser counties information to investigate the systematic differences in the publishing patterns of scientists influenced by the presence or absence of these MDPs.

We find that MDP entries lead to a substantial increase in the production of research relevant to the MDPs and in the commercialization by academic scientists, without a reduction in the scientists' productivity or the quality of their work. These changes are not mainly due to direct collaboration with or funding from the MDPs. Instead, they indicate a shift by scientists toward more applied research, driven by increasing attention and an inherent motivation to acquire future resources and collaboration opportunities presented by the MDPs. The impact of MDP entry exists only in new MDPs and is most pronounced among scientists who are new to topics relevant to the MDP and those without previous patent filings. This supports the notion that there is a shift in attention toward more applied, MDP-relevant problems following an MDP's entry, spurred by the increasing visibility and feasibility of funding and collaboration opportunities from MDPs.

We contribute to several bodies of literature. First, there is increasing interest in understanding the consequences of the interplay between universities and industry on academic knowledge production(Azoulay, Ding, and Stuart 2009; Bikard, Vakili, and Teodoridis 2019; Fini, Perkmann, and Ross 2022; Murray and Stern 2007; Roche 2023). This paper investigates how industry itself shapes academic science and provides large-scale causal evidence across different fields. We also contribute to the understanding of the dynamics of knowledge production and the determinants behind scientists' selection of research topics (Fini, Perkmann, and Ross 2022; Fry 2023; Koning, Samila, and Ferguson 2021; Sohn 2021; Truffa and Wong 2022). While scientists are traditionally viewed as driven by a pure passion for science, accumulating evidence suggests they are also responsive to pecuniary incentives (Agarwal and Ohyama 2013; Sauermann and Roach 2014; Stern 2004). Our study highlights that scientists not only notice but also pay attention to funding and collaboration opportunities from industry. More broadly, we contribute to the literature on how large firms affect innovation. The rising concentration of economic activities in the hands of these large firms raises significant concerns about its potential effects on the innovation process (Bloom et al. 2019; Cunningham, Ederer, and Ma 2021; Gofman and Jin 2022; Greenstone, Hornbeck, and Moretti 2010; Polidoro Jr and Yang 2021). Our study adds a new dimension to this conversation by focusing on the role of universities within this ecosystem.

Our paper also has implications for government officials, university administrators, and firm managers. Local governments in the U.S. have competed fiercely to attract large high-skilled firms in recent decades, hoping that these firms will create jobs and stimulate economic growth. A recent example of this trend was the intense competition among over 200 counties to become the location for Amazon's second headquarters, where they offered a range of tax incentives and other benefits. However, the impact of these large firms extends beyond the labor market, reaching into another critical domain: the production of academic knowledge. Additionally, governments have been actively promoting the commercialization of university-based technology since the enactment of the Bayh-Dole Act in 1980. To encourage the commercialization of university research, the government has put forth significant efforts, such as the NSF's Innovation Corps. Our research suggests that governments can further motivate the generation of commercially viable and firm-relevant research by funding, collaboration, or redirecting scientists' attention.

For university policymakers, the foundational mission of universities has long been centered around the generation and dissemination of knowledge (Bush 1945), but universities are increasingly including economic development and commercial translation into their core objectives (Sanberg et al. 2014). Our study reveals that these dual missions are not necessarily in conflict, as some existing literature suggests. Instead, industry can lead to the redirection of scientists' focus toward more applied areas, fostering unique and innovative knowledge combinations.

For firm managers, the implications of this shift are multifaceted. On the one hand, scientists are not indifferent to industry; instead, they actively pursue funding and collaboration opportunities provided by industry. Firms can take advantage of scientists' attention to firm-relevant topics and integrate them into the R&D process. On the other hand, firms have leveraged academic science to stay updated with new fundamental scientific advancements. However, our findings indicate a potential challenge for firms relying on academic research. When academic scientists pivot their research focus toward areas already familiar to the exploiting firm, it may diminish the anticipated benefits of such exploitation. Therefore, firm managers should consider strategies that not only exploit existing academic research but also actively encourage and support academic inquiries into new, unexplored fields. This approach can ensure a continuous inflow of novel ideas and technologies, keeping firms at the forefront of innovation.

Our findings are not without limitations and therefore open the door to future research. First, our empirical strategy relies on the reported location rankings of profit-maximizing firms, and we assume that loser counties are valid counterfactuals to winner counties. Still, the firms' and scientists' locations are not randomly assigned in our data; hence, it is reasonable to assume that there may still be some systematic differences between the scientists in the winner counties and those in the loser counties. As mentioned, our theory assumes that scientists shift their direction toward firm-relevant fields as they directly receive funding from firms, collaborate with firms, or pay more attention to the funding and collaboration opportunities enabled by firms. While we provide direct evidence supporting our theory and exclude several potential explanations, we cannot deny the possibility of some other mechanisms, such as informal interactions between firms and scientists. Future research should investigate whether and which types of informal interactions are most effective in steering applied knowledge production in academia.

In summary, our research reveals that industry can shape academic science by redirecting academic scientists' research focus toward more applied and firm-relevant subjects. This shift is prompted not only by direct firm investments such as funding and collaboration but also by a deliberate reallocation of scientists' attention toward private, for-profit entities, in anticipation of future resources and collaboration opportunities. This finding has important implications for the symbiotic relationship between industry and academia, suggesting that academic scientists are not reluctant but willing to explore more diverse, remote and applied areas of research when industry opportunities are visible and feasible. We anticipate that our findings on the complex interplay between local firms and academic scientists' knowledge production will serve as a catalyst for further studies on how industry shapes academic scientists' trajectories and how scientists choose their research topics.

References

Agarwal, Rajshree, and Atsushi Ohyama. 2013. "Industry or academia, basic or applied? Career choices and earnings trajectories of scientists." *Management Science* 59 (4): 950–970.

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

Agrawal, Ajay, and Rebecca Henderson. 2002. "Putting patents in context: Exploring knowledge transfer from MIT." *Management science* 48 (1): 44–60.

Ahmadpoor, Mohammad, and Benjamin F Jones. 2017. "The dual frontier: Patented inventions and prior scientific advance." *Science* 357 (6351): 583–587.

Andrews, Michael. 2019. "Bar talk: Informal social interactions, alcohol prohibition, and invention." *Working Paper.*

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

Arora, Ashish, Sharon Belenzon, and Lia Sheer. 2021. "Knowledge spillovers and corporate investment in scientific research." *American Economic Review* 111 (3): 871–898.

Atkin, David, M Keith Chen, and Anton Popov. 2022. *The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley.* Technical report. National Bureau of Economic Research.

Azoulay, Pierre, Waverly 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.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin. 2019. "Does science advance one funeral at a time?" *American Economic Review* 109 (8): 2889–2920.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda. 2020. "Age and high-growth entrepreneurship." *American Economic Review: Insights* 2 (1): 65–82.

Babina, Tania, Alex Xi He, Sabrina T Howell, Elisabeth Ruth Perlman, and Joseph Staudt. 2023. "Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications." *The Quarterly Journal of Economics* 138 (2): 895–954.

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

Bikard, Michaël, Keyvan Vakili, and Florenta Teodoridis. 2019. "When collaboration bridges institutions: The impact of university–industry collaboration on academic productivity." *Organization Science* 30 (2): 426–445.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What drives differences in management practices?" *American Economic Review* 109 (5): 1648–1683.

Bush, Vannevar. 1945. *Science, the endless frontier.* Princeton University Press.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaulé. 2020. "How do travel costs shape collaboration?" *Management Science* 66 (8): 3340–3360.

Chai, Sen. 2017. "Near misses in the breakthrough discovery process." *Organization Science* 28 (3): 411–428.

Cockburn, Iain M, and Rebecca M Henderson. 1998. "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery." *The journal of industrial economics* 46 (2): 157–182.

Cohen, Wesley M, Richard R Nelson, and John P Walsh. 2002. "Links and impacts: the influence of public research on industrial R&D." *Management science* 48 (1): 1–23.

Cunningham, Colleen, Florian Ederer, and Song Ma. 2021. "Killer acquisitions." *Journal of Political Economy* 129 (3): 649–702.

Czarnitzki, Dirk, Christoph Grimpe, and Andrew A Toole. 2015. "Delay and secrecy: Does industry sponsorship jeopardize disclosure of academic research?" *Industrial and Corporate change* 24 (1): 251–279.

Dasgupta, Partha, and Paul A David. 1994. "Toward a new economics of science." *Research policy* 23 (5): 487–521.

Etzkowitz, Henry, Andrew Webster, and Peter Healey. 1998. *Capitalizing knowledge: New intersections of industry and academia.* suny Press.

Evans, James A. 2010. "Industry induces academic science to know less about more." *American Journal of Sociology* 116 (2): 389–452.

Fabrizio, Kira R, and Alberto Di Minin. 2008. "Commercializing the laboratory: Faculty patenting and the open science environment." *Research policy* 37 (5): 914–931.

Fini, Riccardo, Markus Perkmann, and Jan-Michael Ross. 2022. "Attention to exploration: The effect of academic entrepreneurship on the production of scientific knowledge." *Organization Science* 33 (2): 688–715.

Fleming, Lee, Hillary Greene, G Li, Matt Marx, and Dennis Yao. 2019. "Government-funded research increasingly fuels innovation." *Science* 364 (6446): 1139–1141.

Fleming, Lee, and Olav Sorenson. 2004. "Science as a map in technological search." *Strategic management journal* 25 (8-9): 909–928.

Fry, Caroline Viola. 2023. "Crisis and the trajectory of science: Evidence from the 2014 Ebola outbreak." *Review of Economics and Statistics* 105 (4): 1028–1038.

Gofman, Michael, and Zhao Jin. 2022. "Artificial Intelligence, Education, and Entrepreneurship." *Journal of Finance, Forthcoming.*

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." *Journal of Political Economy* 118 (3): 536–598.

Hasan, Sharique, and Rembrand Koning. 2019. "Conversations and idea generation: Evidence from a field experiment." *Research Policy* 48 (9): 103811.

Hausman, Naomi. 2022. "University innovation and local economic growth." *Review of Economics and Statistics* 104 (4): 718–735.

Hegde, Deepak, and Justin Tumlinson. 2014. "Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in US venture capital." *Management Science* 60 (9): 2355–2380.

Hofstra, Bas, Vivek V Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A McFarland. 2020. "The diversity–innovation paradox in science." *Proceedings of the National Academy of Sciences* 117 (17): 9284–9291.

Hsu, David H, and Jeffrey M Kuhn. 2023. "Academic stars and licensing experience in university technology commercialization." *Strategic Management Journal* 44 (3): 887–905.

Jensen, Richard, Jerry Thursby, and Marie C Thursby. 2010. *University-industry spillovers, government funding, and industrial consulting.* Technical report. National Bureau of Economic Research.

Jones, Benjamin F. 2009. "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?" *The Review of Economic Studies* 76 (1): 283–317.

Kafouros, Mario, Chengqi Wang, Panagiotis Piperopoulos, and Mingshen Zhang. 2015. "Academic collaborations and firm innovation performance in China: The role of region-specific institutions." *Research Policy* 44 (3): 803–817.

Kaiser, Ulrich, Hans C Kongsted, Keld Laursen, and Ann-Kathrine Ejsing. 2018. "Experience matters: The role of academic scientist mobility for industrial innovation." *Strategic Management Journal* 39 (7): 1935–1958.

Koffi, Marlène, and Matt Marx. 2023. *Cassatts in the Attic.* Technical report. National Bureau of Economic Research.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson. 2021. "Who do we invent for? Patents by women focus more on women's health, but few women get to invent." *Science* 372 (6548): 1345–1348.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions.* University of Chicago press.

Leahey, Erin, Christine M Beckman, and Taryn L Stanko. 2017. "Prominent but less productive: The impact of interdisciplinarity on scientists' research." *Administrative Science Quarterly* 62 (1): 105–139.

March, James G. 1991. "Exploration and exploitation in organizational learning." *Organization science* 2 (1): 71–87.

Marx, Matt, and Aaron Fuegi. 2020. "Reliance on science: Worldwide front-page patent citations to scientific articles." *Strategic Management Journal* 41 (9): 1572–1594.

———. 2022. "Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations." *Journal of Economics & Management Strategy* 31 (2): 369–392.

Marx, Matt, and David H Hsu. 2022. "Revisiting the entrepreneurial commercialization of academic science: Evidence from "Twin" discoveries." *Management Science* 68 (2): 1330–1352.

Merton, Robert K. 1973. *The sociology of science: Theoretical and empirical investigations.* University of Chicago press.

Mindruta, Denisa. 2013. "Value creation in university-firm research collaborations: A matching approach." *Strategic management journal* 34 (6): 644–665.

Mowery, David C, and Arvids A Ziedonis. 2015. "Markets versus spillovers in outflows of university research." *Research Policy* 44 (1): 50–66.

Murray, Fiona. 2010. "The oncomouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions." *American Journal of sociology* 116 (2): 341–388.

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): 648–687.

Owen-Smith, Jason. 2003. "From separate systems to a hybrid order: Accumulative advantage across public and private science at research one universities." *Research policy* 32 (6): 1081–1104.

Polidoro Jr, Francisco, Curba Morris Lampert, and Minyoung Kim. 2022. "External knowledge sourcing, knowledge spillovers, and internal collaboration: The effects of intrafirm linkages on firm-university co-authorship linkages." *Strategic Management Journal* 43 (13): 2742–2776.

Polidoro Jr, Francisco, and Wei Yang. 2021. "Corporate investment relationships and the search for innovations: An examination of startups' search shift toward incumbents." *Organization Science* 32 (4): 909–939.

Priem, Jason, Heather Piwowar, and Richard Orr. 2022. "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts." *arXiv preprint arXiv:2205.01833.*

Qian, Franklin, and Rose Tan. 2021. "The effects of high-skilled firm entry on incumbent residents." *Stanford Institute for Economic Policy Research (SIEPR) Working Paper,* 21–039.

Roach, Michael, and Wesley M Cohen. 2013. "Lens or prism? Patent citations as a measure of knowledge flows from public research." *Management Science* 59 (2): 504–525.

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): 422–434.

Roche, Maria P. 2023. "Academic entrepreneurship: Entrepreneurial advisors and their advisees' outcomes." *Organization Science* 34 (2): 959–986.

Roche, Maria P, Alexander Oettl, and Christian Catalini. 2023. "Proximate (Co-) Working: Knowledge Spillovers and Social Interactions."

Rosenberg, Nathan. 1982. *Inside the black box: technology and economics.* cambridge university press.

Rosenberg, Nathan, and Richard R Nelson. 1994. "American universities and technical advance in industry." *Research policy* 23 (3): 323–348.

Rosenberg, Nathan, Scott Stern, Eunhee Sohn, and Michael J. Andrews. 2024. "The Endogeneity of Science." *Working Paper.*

Sanberg, Paul R, Morteza Gharib, Patrick T Harker, Eric W Kaler, Richard B Marchase, Timothy D Sands, Nasser Arshadi, and Sudeep Sarkar. 2014. "Changing the academic culture: Valuing patents and commercialization toward tenure and career advancement." *Proceedings of the National Academy of Sciences* 111 (18): 6542–6547.

Sauermann, Henry, and Michael Roach. 2014. "Not all scientists pay to be scientists: PhDs' preferences for publishing in industrial employment." *Research Policy* 43 (1): 32–47.

Sauermann, Henry, and Paula Stephan. 2013. "Conflicting logics? A multidimensional view of industrial and academic science." *Organization science* 24 (3): 889–909.

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* 77 (5): 804–830.

Singh, Jasjit, and Ajay Agrawal. 2011. "Recruiting for ideas: How firms exploit the prior inventions of new hires." *Management science* 57 (1): 129–150.

Sohn, Eunhee. 2021. "How local industry R&D shapes academic research: Evidence from the agricultural biotechnology revolution." *Organization Science* 32 (3): 675–707.

Stephan, Paula. 1996. "The economics of science." *Journal of Economic literature* 34 (3): 1199–1235.

———. 2012. *How economics shapes science.* Harvard University Press.

Stephan, Paula E, Shiferaw Gurmu, Albert J Sumell, and Grant Black. 2007. "Who's patenting in the university? Evidence from the survey of doctorate recipients." *Econ. Innov. New Techn.* 16 (2): 71–99.

Stern, Scott. 2004. "Do scientists pay to be scientists?" *Management science* 50 (6): 835–853.

Stokes, Donald E. 2011. *Pasteur's quadrant: Basic science and technological innovation.* Brookings Institution Press.

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.

Toole, Andrew A, and Dirk Czarnitzki. 2010. "Commercializing science: Is there a university "brain drain" from academic entrepreneurship?" *Management science* 56 (9): 1599–1614.

Truffa, Francesca, and Ashley Wong. 2022. *Undergraduate gender diversity and direction of scientific research.*

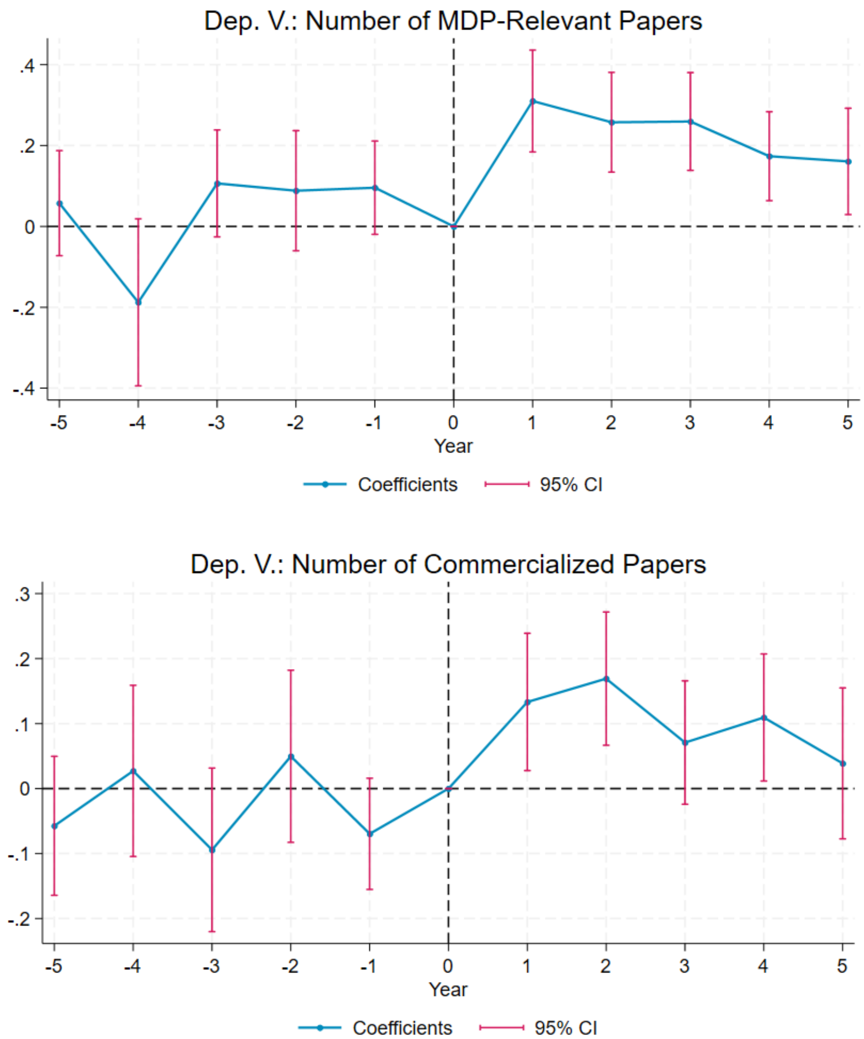
Tzabbar, Daniel. 2009. "When does scientist recruitment affect technological repositioning?" *Academy of Management Journal* 52 (5): 873–896.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical combinations and scientific impact." *Science* 342 (6157): 468–472.

Zucker, Lynne G, Michael R Darby, and Jeff Armstrong. 1998. "Geographically localized knowledge: spillovers or markets?" *Economic inquiry* 36 (1): 65–86.

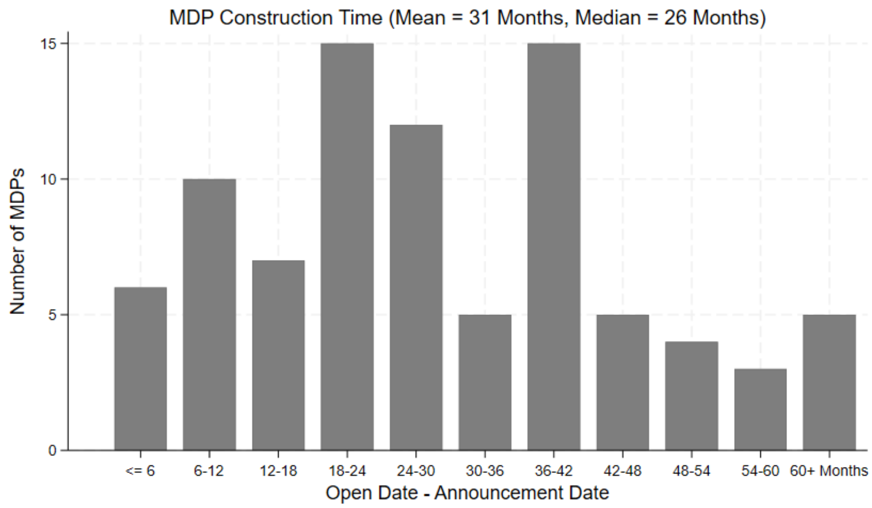
Figures

Figure 1: Event Study - Yearly Treatment Effects of MDPs



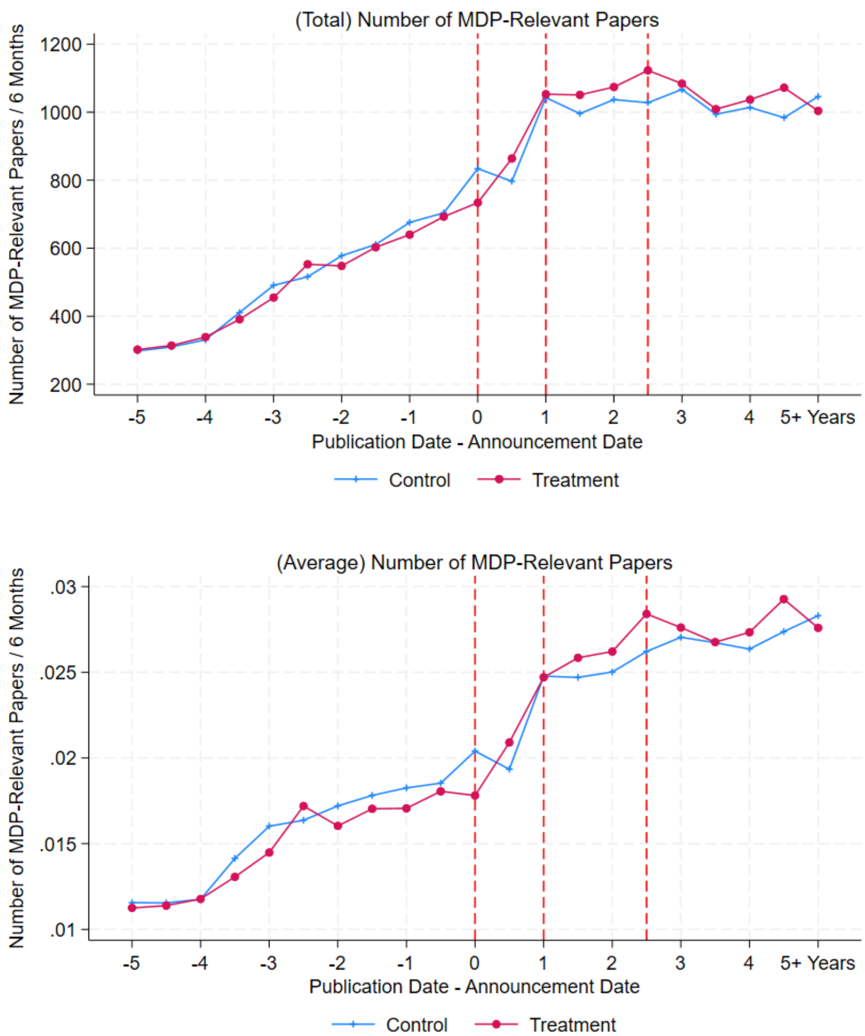
Notes. The blue solid line corresponds to the coefficient estimates from conditional fixed-effects quasi-maximum likelihood Poisson regressions in which the number of MDP-Relevant papers and the number of commercialized papers are regressed on the interaction terms between the treatment status and the number of years before/after treatment event (the announcement of MDPs). *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the MDP. *Number of Commercialized Papers* is the number of papers published in year t that are commercialized in a patent-paper pair by scientist i. Pair, scientist, and year fixed effects are included. Robust standard errors were clustered at the pair and scientist level. The light red lines show the 95% confidence intervals around these estimates.

Figure 2: Construction Time: From Announcement to Opening



Notes. The figure shows the distribution of time gaps between MDP announcement time and MDP opening time.

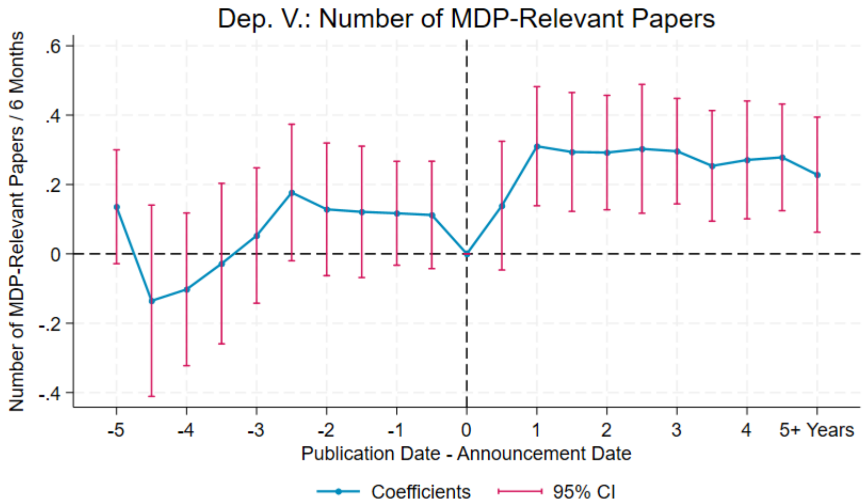
Figure 3: Summary Statistics: MDP-Relevant Papers



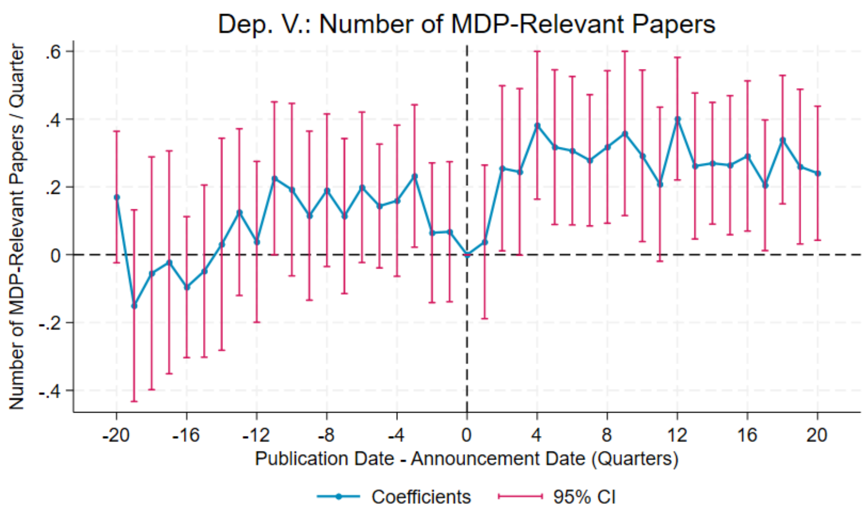
Notes. The figure shows the aggregate number of MDP-relevant papers of treatment and control counties, and the average number of MDP-relevant papers of scientists in treatment and control counties.

Figure 4: Event Study - Semiannual and Quarterly Treatment Effects of MDPs

(a) Semiannual Effects

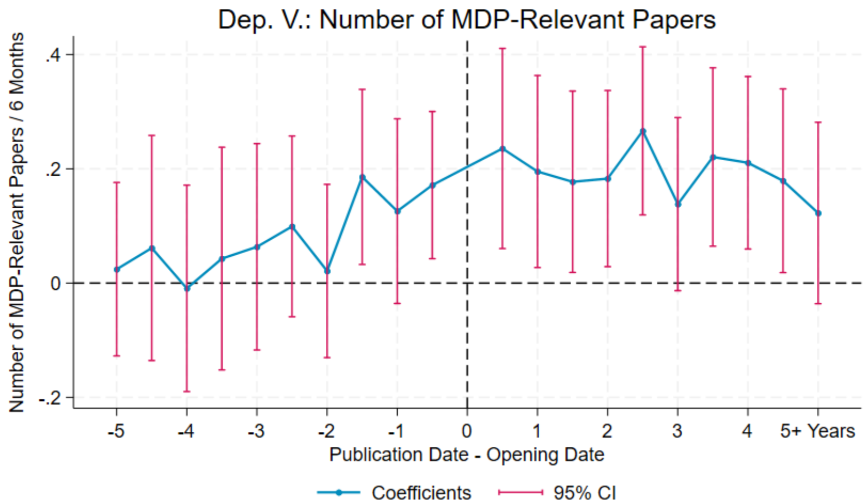


(b) Quarterly Effects



Notes. The blue solid line corresponds to the coefficient estimates from conditional fixed-effects quasi-maximum likelihood Poisson regressions in which the number of MDP-Relevant papers is regressed on the interaction terms between the treatment status and the number of half-years or quarters before/after treatment event (the announcement of MDPs). The time difference is calculated based on publication date and announcement date, instead of calendar year. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the MDP. Pair, scientist, and year fixed effects are included. Robust standard errors were clustered at the pair and scientist level. The light red lines show the 95% confidence intervals around these estimates.

Figure 5: Placebo Test: Opening Date



Notes. The blue solid line corresponds to the coefficient estimates from conditional fixed-effects quasi-maximum likelihood Poisson regressions in which the number of MDP-Relevant papers and number of patents are regressed on the interaction terms between the treatment status and the number of half-years before/after treatment event (the opening of MDPs). *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the MDP. Pair, scientist, and year fixed effects are included. Robust standard errors were clustered at the pair and scientist level. The light red lines show the 95% confidence intervals around these estimates.

Tables

Table 1: Descriptive Statistics

Var.	N	Mean	S.D.	Min	Max
Panel A: Scientist-Year					
Number of Papers	1817585	3.1346	15.3015	1.00	14010.00
Paper Citations	1817585	11.2570	37.4877	0.00	21101.00
Number of MDP-Relevant Papers	1817585	0.0327	0.2421	0.00	27.00
Number of Commercialized Papers	1817585	0.0147	0.1417	0.00	9.00
Number of Pat-Cited Papers	1817585	0.0969	0.3713	0.00	20.00
Patent Citations	1817585	0.1036	0.8906	0.00	235.00
Number of MDP-Funded Papers	1817585	0.0011	0.0359	0.00	5.00
Number of MDP-Coauthored Papers	1817585	0.0001	0.0130	0.00	5.00
Number of Different Concepts	1817585	5.7306	6.1770	0.00	1514.00
Academic Age at MDP Entry	1817585	17.4412	12.9262	0.00	60.00
Never Worked on MDP-Relevant Topics	1817585	0.9136	0.2810	0.00	1.00
Rarely Worked on MDP-Relevant Topics	1817585	0.9888	0.1052	0.00	1.00
Never Commercialized Papers	1817585	0.9528	0.2122	0.00	1.00
Rarely Coauthored w. Firms	1817585	0.9313	0.2530	0.00	1.00
Panel B: University-Year					
R1 or Not	13117	0.1192	0.3240	0.00	1.00
Number of Earned Doctorates / 1000	13117	0.0342	0.1088	0.00	0.92
R&D Grants / 1000	13117	29.7824	116.6325	0.00	1675.81
R&D Grants from Corp. / 1000	13117	1.9347	10.8227	0.00	252.37
R&D Grants from NSF / 1000	13117	2.5041	10.0567	0.00	142.61
R&D Grants from NIH / 1000	13117	8.8499	40.4616	0.00	656.34
Number of Papers / 1000	13117	0.3892	3.5448	0.00	386.07
Number of Pat-cited Papers / 1000	13117	0.0188	0.0672	0.00	0.84
Number of MDP-Relevant Papers / 1000	13117	0.0060	0.0268	0.00	0.54
Panel C: MDP					
Announcement Year	105	2007	4.8284	2000	2015
MDP: Low Skill or Not	105	0.1000	0.3015	0	1
MDP: New or Expansion	105	0.7000	0.4606	0	1
MDP: Fund Research or Not	105	0.5300	0.5016	0	1

Notes. Each observation in Panel A is at the scientist-year level. *Number of Papers* is the number of publications of scientist i published in year t. *Paper Citations* is the mean of the 5-year forward citations of papers published in year t. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Number of Commercialized papers* is the number of papers published in year t that are commercialized in a patent-paper pair. *Number of Pat-Cited Papers* is the number of papers published in year t that are cited by patents within 5 years. *Patent Citations* is the mean of the 5-year patent citation of papers published in year t. Variables in Panel B are at the institutions level (R1 or not, total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers). Panel C summarizes MDP statistics (announcement year, low skill or not, new or expansion, fund research or not).

Table 2: Correlations

Var.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Number of Papers	1.00													
Paper Citations	0.00	1.00												
Number of MDP-Relevant Papers	0.06	0.00	1.00											
Number of Commercialized Papers	0.02	0.04	0.06	1.00										
Number of Pat-Cited Papers	0.04	0.13	0.09	0.30	1.00									
Patent Citations	-0.00	0.16	0.01	0.14	0.36	1.00								
Number of MDP-Funded Papers	0.01	0.01	0.00	0.01	0.03	0.01	1.00							
Number of MDP-Coauthored Papers	0.00	0.00	0.01	0.01	0.01	0.00	0.00	1.00						
Number of Different Concepts	0.45	0.02	0.14	0.07	0.18	-0.01	0.05	0.00	1.00					
Academic Age at MDP Entry	0.03	-0.04	-0.02	-0.03	-0.04	-0.04	-0.00	-0.00	0.08	1.00				
Never Worked on MDP-Relevant Topics	0.02	-0.01	0.28	0.02	0.03	0.00	-0.00	0.01	0.08	-0.06	1.00			
Rarely Worked on MDP-Relevant Topics	0.01	-0.00	0.24	0.02	0.03	0.01	-0.00	0.00	0.05	-0.03	0.35	1.00		
Never Commercialized Papers	0.01	0.02	0.03	0.23	0.14	0.07	0.00	0.01	0.05	-0.03	0.03	0.03	1.00	
Rarely Coauthored w. Firms	0.05	0.01	0.02	0.03	0.07	0.01	0.02	0.00	0.20	-0.13	0.06	0.05	0.12	1.00

Notes. Each observation is at the scientist-year level.

Table 3: Regression Analysis of T. and C. Scientists' Pre-Treatment Characteristics

	(1)	(2)	(3)	(4)
	Quantity and Quality		MDP Relevance	Commercialization
D.V.	Number of Papers	Paper Citations	Number of MDP-Relevant Papers	Number of Commercialized Papers
Treat	0.00250	0.01917	0.13358	0.13401
	(0.00768)	(0.07178)	(0.11477)	(0.10731)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	Poisson	OLS	Poisson	Poisson
Observations	871447	871447	104544	51033

Notes. Each observation is at the scientist-year level and is pre-treatment observation. Models (1), (3) and (4) report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Model (2) reports OLS regression coefficients with robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of Papers* is the number of publications of scientist i published in year t. *Paper Citations* is the mean of the 5-year forward citations of papers published in year t. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Number of Commercialized Papers* is the number of papers published in year t that are commercialized in a patent-paper pair by scientist i. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 4: The Impact of MDPs on Scientists' Research Quantity and Quality

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	Number of Papers			Paper Citations		
Treat × Post	-0.00894	-0.00893	-0.00493	0.56426***	0.59221***	0.60109***
	(0.02084)	(0.01682)	(0.01689)	(0.06121)	(0.14702)	(0.12243)
Pair FE	No	Yes	Yes	No	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Model	Poisson	Poisson	Poisson	OLS	OLS	OLS
Observations	1816121	1816121	1816121	1816121	1816121	1816121

Notes. Each observation is at the scientist-year level. Models (1) to (3) report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Models (4) to (6) report OLS regression coefficients with robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. The dependent variable (D.V.) for models (1) to (3) is the number of publications of scientist i published in year t. The dependent variable (D.V.) for models (4) to (6) is the mean of the 5-year forward citations of papers published in year t. Models (3) and (6) include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 5: The Impact of MDPs on Scientists' Research Direction and Commercialization

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	Number of MDP-Relevant Papers			Number of Commercialized Papers		
Treat × Post	0.11018***	0.17768***	0.17935***	0.10347***	0.11744***	0.11994***
	(0.02088)	(0.05204)	(0.04971)	(0.02141)	(0.03349)	(0.03745)
Pair FE	No	Yes	Yes	No	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Model	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Observations	423578	347045	347045	185231	185231	185231

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Number of Commercialized Papers* is the number of papers published in year t that are commercialized in a patent-paper pair by scientist i. Models (3) and (6) include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 6: Mechanism Test: MDP - Skills and Funding

	(1)	(2)	(3)	(4)
D.V.	Number of MDP-Relevant Papers			
	High-Skilled MDPs	Low-Skilled MDPs	MDPs Fund Res.	MDPs Don't Fund Res.
Treat × Post	0.18553***	-0.10248	0.16008***	0.11572
	(0.05056)	(0.17777)	(0.04733)	(0.13235)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson	Poisson
Observations	323773	14708	296664	29975

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Low-skill* and *High-skilled* MDPs are defined as whether the MDP hires a high-skilled labor force. *MDPs Fund Research* and *MDPs Don't Fund research* are defined as whether the parent firm of the entered MDP actively funds research. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 7: Mechanism Test: Funding, Collaboration or Not

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
D.V.	Number of			Number of Papers	Number of MDP-Relevant Papers		
	MDP-Funded Papers	MDP-Coauthored Papers			Exclude		
					MDP Coauthors	MDP Grantees	Both
Treat × Post	0.19399*	-0.98879**	1.34316*	0.00866	0.17930***	0.18042***	0.18018***
	(0.11032)	(0.46286)	(0.76534)	(0.00593)	(0.04945)	(0.05012)	(0.04986)
˜ × (CoauthorWithMDP)			-3.83352***	0.10578*			
			(0.82567)	(0.06073)			
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Observations	13373	1737	1737	1816121	345962	342438	341375

Notes. Each observation is at the scientist-year level. All Models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of MDP-Funded Papers* is the number of papers published in year t that are funded by the MDP. *Number of MDP-Coauthored Papers* is the number of papers published in year t that are coauthored with the MDP. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 8: Mechanism Test: New Entry or Expansion

	(1)	(2)
D.V.	Number of MDP-Relevant Papers	
	New MDP	Expansion
Treat × Post	0.16235***	0.11186
	(0.03586)	(0.09882)
Pair FE	Yes	Yes
Scientist FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes
Model	Poisson	Poisson
Observations	240584	85977

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *New MDP* indicates that the MDP entered the county for the first time instead of an *expansion* of the existing MDP. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table 9: Mechanism Test: Prior Experience

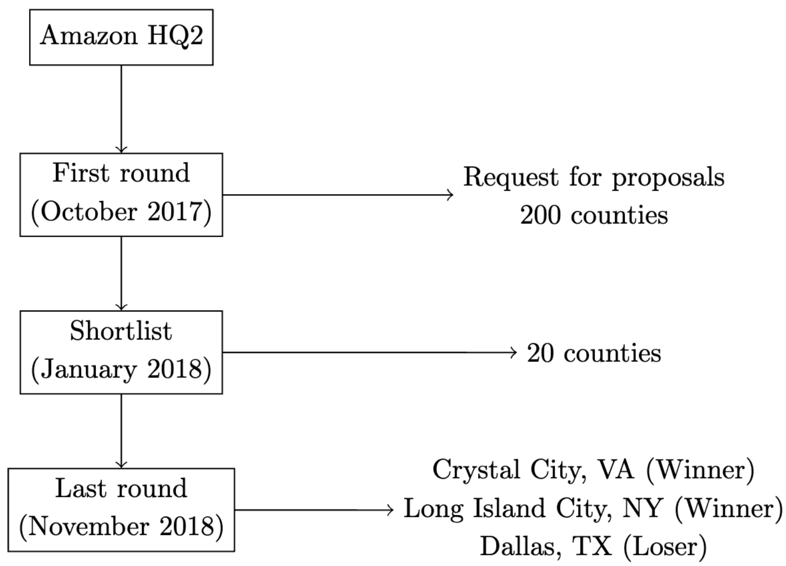
	(1)	(2)	(3)	(4)
D.V.	Number of MDP-Relevant Papers			
Treat × Post	-0.25624***	-0.26885***	0.06934	0.01042
	(0.07203)	(0.06573)	(0.13820)	(0.10179)
˜ × (NeverWorkMDPRelevant)	1.22540***			
	(0.23684)			
˜ × (RarelyWorkMDPRelevant)		0.55738***		
		(0.06902)		
˜ × (NeverCommercializePapers)			0.12120	
			(0.11270)	
˜ × (RarelyCoauthorWithFirms)				0.19042**
				(0.07709)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson	Poisson
Observations	347045	347045	347045	347045

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *NeverWorkMDPRelevant* equals one if the scientist has never worked on MDP-relevant concepts before the entry of MDPs. *RarelyWorkMDPRelevant* equals one if the scientist has less than three publications that are MDP-relevant before the entry of MDPs. *NeverCommercializePapers* indicates whether the scientist filed patents before. *RarelyCoauthorWithFirms* is a binary variable switches to one when the scientist has less than three papers coauthored with any firms previously. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

A Appendix

A.1 Site Selection of MDPs

Figure A.1: Site Selection of Amazon HQ2



A.2 MDP-Relevant Papers

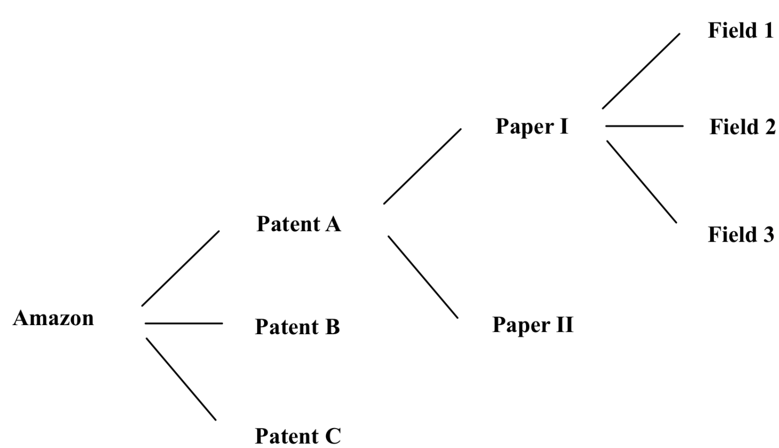
We define whether a paper is MDP-relevant based on the following steps:

- Suppose Amazon is planning to open a MDP, for example HQ2, somewhere.

1. We obtain Amazon's public patent records from the USPTO;
2. We merge the patents with the patent-to-paper citations data based on the patent ID. The patent-to-paper citations data is constructed by Marx and Fuegi 2020, 2022, and can be downloaded at https://relianceonscience.org/;
3. We assemble each paper's concepts using the concepts information provided by OpenAlex. The concepts are abstract ideas that the papers are about, details about the definition of concepts can be found at https://docs.openalex.org/api-entities/concepts. We only use layer-2 concepts of all the publications to ensure consistency and granularity.
4. We calculate Amazon's top 20 most cited concepts, and define papers containing these concepts as Amazon-relevant papers.

- For scientists affected by different MDPs, their MDP-relevant papers correspond to the entered MDPs.

Figure A.2: MDP-Relevant Papers (Example)

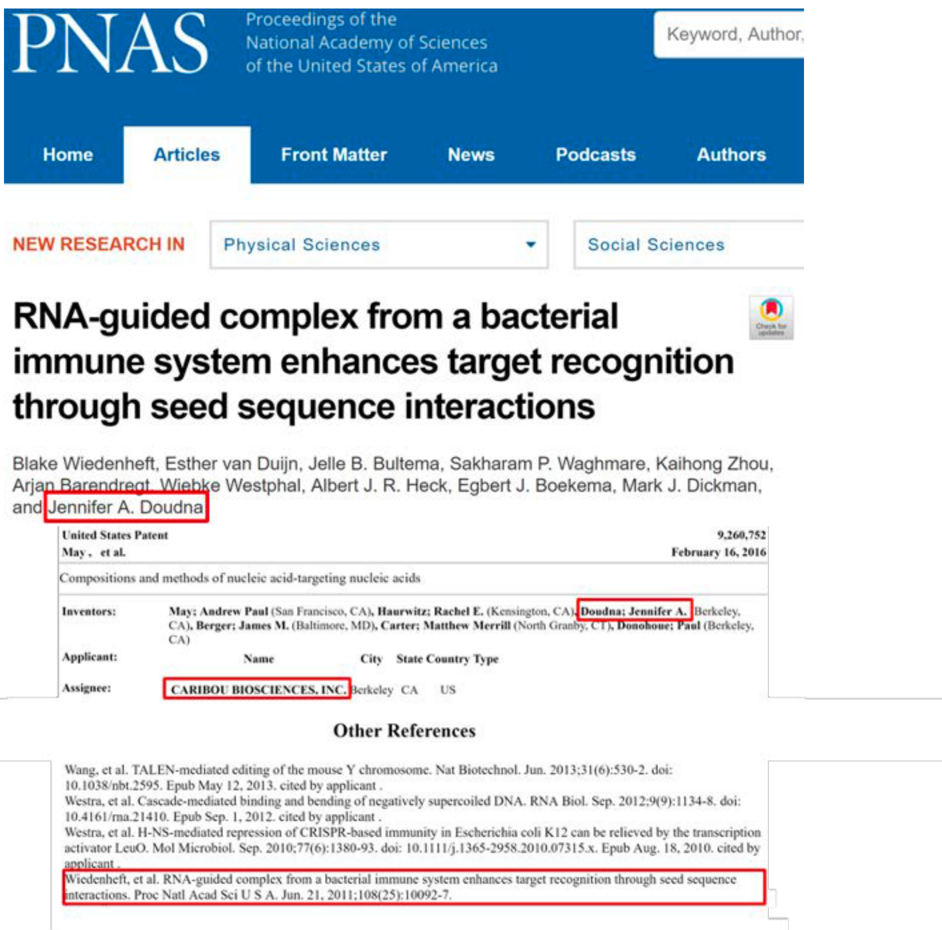


A.3 Commercialized Papers

Commercialized papers are defined based on patent-paper pairs data, the data can be downloaded at https://relianceonscience.org/. A paper has a patent-paper pair if:

- The paper is cited by a patent;
- The patent assignee is a firm;
- The inventors on the patent overlap with the authors of the paper.

Figure A.3: Commercialized Paper (Example)



PNAS
Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author,

Home | Articles | Front Matter | News | Podcasts | Authors

NEW RESEARCH IN Physical Sciences | Social Sciences

RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions

Blake Wiedenheft, Esther van Duijn, Jelle B. Bultema, Sakharam P. Waghmare, Kaihong Zhou, Arjan Barendregt, Wiebke Westphal, Albert J. R. Heck, Egbert J. Boekema, Mark J. Dickman, and Jennifer A. Doudna

United States Patent **9,260,752**
May , et al. **February 16, 2016**

Compositions and methods of nucleic acid-targeting nucleic acids

Inventors: **May; Andrew Paul** (San Francisco, CA), **Haurwitz; Rachel E.** (Kensington, CA), **Doudna; Jennifer A.** Berkeley, CA), **Berger; James M.** (Baltimore, MD), **Carter; Matthew Merrill** (North Granby, CT), **Donohoue; Paul** (Berkeley, CA)

Applicant: **Name** **City** **State Country Type**

Assignee: **CARIBOU BIOSCIENCES, INC.** Berkeley CA US

Other References

Wang, et al. TALEN-mediated editing of the mouse Y chromosome. Nat Biotechnol. Jun. 2013;31(6):530-2. doi: 10.1038/nbt.2595. Epub May 12, 2013. cited by applicant .
Westra, et al. Cascade-mediated binding and bending of negatively supercoiled DNA. RNA Biol. Sep. 2012;9(9):1134-8. doi: 10.4161/rna.21410. Epub Sep. 1, 2012. cited by applicant .
Westra, et al. H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol. Sep. 2010;77(6):1380-93. doi: 10.1111/j.1365-2958.2010.07315.x. Epub Aug. 18, 2010. cited by applicant .
Wiedenheft, et al. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc Natl Acad Sci U S A. Jun. 21, 2011;108(25):10092-7.

B Appendix: Tables

Table B.1: t-Test of Treated and Control Scientists' Pre-Treatment Characteristics: CEM Sample

Var.	Observations	Mean	SD	Observations	Mean	SD	Diff	p-value
Number of Papers	451259	2.730	3.95	446175	2.727	3.27	0.003	(0.74)
Paper Citations	451259	10.445	16.78	446175	10.304	16.89	0.141	(0.00)
Number of MDP-Relevant Papers	451259	0.022	0.17	446175	0.023	0.18	-0.001	(0.01)
Number of Commercialized Papers	451259	0.010	0.11	446175	0.011	0.12	-0.001	(0.00)
Number of Pat-Cited Papers	451259	0.088	0.31	446175	0.089	0.31	-0.001	(0.15)
Patent Citations	451259	0.098	0.59	446175	0.101	0.61	-0.003	(0.01)
Number of MDP-Funded Papers	451259	0.001	0.03	446175	0.001	0.03	0.000	(0.15)
Number of MDP-Coauthored Papers	451259	0.000	0.01	446175	0.000	0.01	0.000	(0.00)
Number of Different Concepts	451259	5.097	4.90	446175	5.100	4.92	-0.003	(0.78)
Academic Age at MDP Entry	451259	18.998	12.77	446175	19.057	12.90	-0.058	(0.03)
Never Worked on MDP-Relevant Topics	451259	0.913	0.28	446175	0.907	0.29	0.006	(0.00)
Never Commercialized Papers	451259	0.951	0.22	446175	0.949	0.22	0.003	(0.00)
Rarely Worked on MDP-Relevant Topics	451259	0.988	0.11	446175	0.988	0.11	0.000	(0.75)
Rarely Coauthored w. Firms	451259	0.923	0.27	446175	0.924	0.26	-0.001	(0.06)

Notes. Each observation is at the scientist-year level and is pre-treatment observation. scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of Papers* is the number of publications of scientist i published in year t. *Paper Citations* is the mean of the 5-year forward citations of papers published in year t. *Number of MDP-relevant papers* is the number of papers published in year t that are relevant to the entered MDP. *Number of Commercialized Papers* is the number of papers published in year t that are commercialized in a patent-paper pair by scientist i. *Number of Pat-Cited Papers* is the number of papers published in year t that are cited by patents within 5 years. *Patent Citations* is the mean of the 5-year patent citation of papers published in year t.

Table B.2: MDP's Decision to Choose Winner Counties

	(1)	(2)
	Winner/Treated County	
Number of Papers	-0.00215	-0.00080
	(0.00174)	(0.00448)
Number of MDP-Relevant Papers	-0.00042	0.03651
	(0.02236)	(0.07968)
Number of Commercialized Papers	0.02280	-0.32853
	(0.14700)	(0.37445)
Pair FE	No	Yes
Decision Year FE	No	Yes
Controls	Yes	Yes
Adjusted R-squared	0.071	0.259
Observations	298	282

Notes. Each observation is at the county level. *Number of Papers* is the number of publications of all scientists in a county published before the entry of MDPs. *Number of MDP-relevant papers* is the number of papers published before the entry of MDPs that are relevant to the entered MDP by all scientists in a county. *Number of Commercialized Papers* is the number of papers published before the entry of MDPs that are commercialized in a patent-paper pair by all scientists in a county. All models include controls for county level docral students, total R&D grants, R&D grants from firms, NSF, and NIH. $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table B.3: Heterogeneity Test

	(1)	(2)	(3)
D.V.	Number of MDP-Relevant Papers		
Treat × Post	0.27669***	0.21797**	0.26733***
	(0.05973)	(0.09214)	(0.05914)
˜ × Age	-0.01258**		
	(0.00546)		
˜ × Age^2	0.00025**		
	(0.00011)		
˜ × Distance		-0.00699	
		(0.00653)	
˜ × R1 University			-0.11105
			(0.08260)
Pair FE	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson
Observations	347045	347045	347045

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Age* is academic age of the scientist at year t, and academic age is calculated as the time length between year t and the year of the scientist's first publication. *Distance* measures the linear distance from the entered MDP to the location of the scientists' institutions. *R1 University* equals one if the scientist is affiliated in an institution classified as *Research I* by Carnegie Classification of Institutions of Higher Education. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table B.4: Robustness Check: Model Specification (OLS)

	(1)	(2)	(3)	(4)
	Quantity and Quality		MDP Relevance	Commercialization
D.V.	Number of Papers	Paper Citations	Number of MDP-Relevant Papers	Number of Commercialized Papers
Treat × Post	-0.04118	0.60109***	0.00483***	0.00205***
	(0.06686)	(0.12243)	(0.00173)	(0.00076)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	OLS	OLS	OLS	OLS
Observations	1816121	1816121	1816121	1816121

Notes. Each observation is at the scientist-year level. All models report OLS regression coefficients with robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of Papers* is the number of publications of scientist i published in year t. *Paper Citations* is the mean of the 5-year forward citations of papers published in year t. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. *Number of Commercialized Papers* is the number of papers published in year t that are commercialized in a patent-paper pair by scientist i. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table B.5: Robustness Check: Model Specification (Extensive Margin)

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	MDP-Relevant Papers or Not			Commercialized Papers or Not		
Treat × Post	0.00158***	0.00289***	0.00298***	0.00144***	0.00140***	0.00142**
	(0.00052)	(0.00047)	(0.00098)	(0.00027)	(0.00046)	(0.00054)
Pair FE	No	Yes	Yes	No	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Model	OLS	OLS	OLS	OLS	OLS	OLS
Observations	1816121	1816121	1816121	1816121	1816121	1816121

Notes. Each observation is at the scientist-year level. All models report OLS regression coefficients with robust standard errors clustered at the pair and scientist level in parentheses. scientists in the *Treatment* group are those who are in universities located in winner counties. *MDP-Relevant Papers or Not* is whether the scientist i has any paper published in year t that is relevant to the entered MDP. *Commercialized Papers or Not* is whether the scientist i has any paper published in year t that is commercialized in a patent-paper pair. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table B.6: Robustness Check: Dependent Variables

	(1)	(2)	(3)	(4)
D.V.	Number of Industry-Relevant Papers	Number of Papers w. Firm Coauthors	Number of Pat-Cited Papers	Patent Citations
Treat × Post	0.03367	0.01054	0.09827***	0.01818***
	(0.02487)	(0.01630)	(0.01637)	(0.00418)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson	Poisson
Observations	522651	853236	752032	1816121

Notes. Each observation is at the scientist-year level. Both models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *Number of Industry-relevant papers* is the number of publications published in year t that are relevant to the MDP's industry. *Number of Papers w. Firm Coauthors* is the number of publications published in year t that has firm (not only the entered MDP) coauthors. *Number of Pat-Cited Papers* is the number of papers published in year t that are cited by patents within 5 years. *Patent Citations* is the mean of the 5-year patent citation of papers published in year t. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.

Table B.7: Robustness Check: Age or Experience

	(1)	(2)	(3)	(4)
D.V.	Number of MDP-Relevant Papers			
Treat × Post	-0.25936***	-0.26501***	0.06993	0.01767
	(0.07201)	(0.06670)	(0.13544)	(0.09592)
˜ × (NeverWorkMDPRelevant)	1.22587***			
	(0.23799)			
˜ × (RarelyWorkMDPRelevant)		0.54945***		
		(0.07121)		
˜ × (NeverCommercializePapers)			0.11513	
			(0.11010)	
˜ × (RarelyCoauthorWithFirms)				0.17721**
				(0.07137)
Pair FE	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Academic Age FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Model	Poisson	Poisson	Poisson	Poisson
Observations	347045	347045	347045	347045

Notes. Each observation is at the scientist-year level. All models report Poisson regression coefficients with QML robust standard errors clustered at the pair and scientist level in parentheses. Scientists in the *Treatment* group are those who are in universities located in winner counties. *NeverWorkMDPRelevant* equals one if the scientist had never worked on MDP-relevant concepts before the entry of MDPs. *RarelyWorkMDPRelevant* equals one if the scientist has less than three publications that are MDP-relevant before the entry of MDPs. *NeverCommercializePapers* indicates whether the scientist filed patents before. *RarelyCoauthorWithFirms* is a binary variable switches to one when the researcher has less than three papers coauthored with any firms before. *Number of MDP-Relevant Papers* is the number of papers published in year t that are relevant to the entered MDP. All models include controls for time-varying characteristics of the scientist (number of papers, number of different concepts), institutions (total R&D grants, R&D grants from firms, NSF, and NIH, number of papers, number of MDP-relevant papers, number of patent-cited papers), and fixed effects (pair, scientist, academic age, and year). $*p < 0.10$; $**p < 0.05$; $***p < 0.01$.