Bringing Science to Market: Knowledge Foundations and Performance*

Justine Boudou
Harvard Business School

Maria Roche
Harvard Business School

February 4, 2024

Abstract

Possessing unique knowledge is widely considered a critical source of competitive advantage. In this paper, we assess whether new ventures that exploit their founders' technologically unique knowledge perform better than those that do not (or only little). Using a panel dataset of 510 academic startups in biomedicine created between 2005 and 2015, we find that, perhaps contrary to expectations, startups relying heavily on their founders' academic work are less likely to be acquired and to receive higher levels of funding, while we do not observe differences in their likelihood to go through an IPO. We show that these results are consistent with these startups using a different *type* of scientific knowledge: more narrow and specialized. Besides their outcomes on the commercial market, we further examine the potential ramifications on the academic market. Applying a difference-in-difference design, we find that founders who build extensively on their academic work as the foundation of their startup experience a decrease in both the number of publications and top publications. Our crude back-of-the-envelope calculations suggest that this decrease represents a potential loss of value generated from publications of 40,000 - 333,000 dollars per year per academic entrepreneur that relies strongly on their own research as the basis of the firm. Additional analyses reveal that these results are consistent with a strategic shift towards intellectual property protection, leading these founders to prioritize the protection and commercialization of their specialized knowledge. Overall, our findings underscore the nuanced interplay of academic founders' decisions, highlighting critical implications for both academic and entrepreneurial pursuits.

Keywords: *Firm Performance*, *Exits*, *Knowledge Foundations*, *Academic Startups*,

*The authors thank Annamaria Conti for data access. Many thanks to participants at the Barcelona School of Economics Entrepreneurship Working Group, and DRUID. We are indebted to Ashish Arora, George Chondrakis, Shane Greenstein, Jacqueline Lane, Brian Silverman, and Keyvan Vakili for feedback on early drafts. We gratefully acknowledge funding from the Harvard Business School Division of Research and Faculty Development. All errors and omissions are our own. Contact: jboudou@hbs.edu; mroche@hbs.edu

1 Introduction

Possessing unique knowledge has long been heralded a critical component in achieving competitive advantage (Barney, 1991). Departing from the vision of perfect competition, where information symmetries exist and every firm can copy what the other is doing, knowledge asymmetries as embodied in unique – often tacit – knowledge can help a firm differentiate. The exploitation of such special information may even provide the basis for a new company to form in the first place (Alvarez and Busenitz, 2001) and have a persistent impact on the performance of a firm (Conti and Roche, 2021; Geroski et al., 2010).

In this paper, we examine the extent to which startups that exploit their founders' technologically unique knowledge perform better than those that do not (or only little) and thereby pay particular attention to the specific type of knowledge these founders incorporate in their ventures. We study this question in the context of academic entrepreneurship, which appears especially suitable due to the predominant role of academia in producing specialized knowledge (National Science Board, 2018), the increasing role of science in the market for technologies (Arora, Belenzon and Suh, 2022), and the importance of specialized knowledge in startup formation within this setting. Long recognized to be among the major drivers of economic growth (Dasgupta and David, 1994), universities have been increasingly expanding their traditional role as producers of knowledge into the commercialization of scientific discoveries (Bhaskarabhatla and Hegde, 2014; Crespi et al., 2011; Hsu et al., 2007; Mowery et al., 2004). Perhaps not surprisingly, most recent work provides evidence suggesting that the use of science emanating from academic labs leads to higher quality inventions (Arora, Belenzon and Suh (2022), Krieger et al. (2022)). Academic spinouts in particular – defined as startups founded by faculty and students around technological knowledge developed within academic laboratories (Shah and Pahnke, 2014) – have been increasing at a stunning rate as a consequence (Grimaldi et al., 2011; Roche et al., 2020).[1]

However, despite large public investment, and broad enthusiasm for such activity (Roach, 2017), academic spinouts do not perform particularly well on the market on average (Roche et al., 2020). Although some studies have highlighted that critical heterogeneity among

[1]Some of the more famous examples of academic spinouts include Tableau (founded by Pat Hanrahan, professor at Stanford), Genentech (launched by Herbert Boyer, professor at the University of California), and Duolingo (initiated by Luis von Ahn, professor at Carnegie Mellon University).

academic spinouts may exist (mainly highlighting the role of outliers, so called "stars") (Zucker et al., 2002, 1998), we have yet to better understand the sources of performance differentials. The one source of potential advantage we seek to examine in this paper is related to academic founders' own unique technological knowledge. Leveraging such expertise acquired over years spent in their occupational training (Colombo and Piva, 2012) as the foundation of their new venture could be an important antecedent for success. Building on established strategy theories, we may expect such access to unique resources to present an important source of competitive advantage on the market.

To test this, we create a panel data set of 510 academic startups in biomedicine created between 2005 and 2015 and use a variety of sources in order to characterize their knowledge base and that of their founders. This sector is particularly well suited for our study because it is tightly linked to academic research and has a relatively high propensity to patent, which enables us to capture the initial knowledge base of a startup. In addition, we have detailed information about each startup regarding its founding date and place, founders, patents and exit events. We further link each academic founder to their research and patenting output using Dimensions AI[2].

Our main variable of interest is the extent to which academic founders rely on their own academic work when creating their startup, which we also conceptualize as the distance between founders' academic work and the initial knowledge base of their startup. Proximity between the two knowledge-bases should indicate the degree to which founders rely more on their specialized knowledge when founding their startup, while a larger distance should capture a more broad and general approach to founding. To examine founders' academic work, we consider their published papers before startup creation. We capture the initial knowledge base of a startup by looking at its first granted patents, at or close to the time of creation. To operationalize the distance variable, we then calculate the percentage of citations that these patents make to their founders' academic papers using the Reliance on Science dataset (Marx and Fuegi, 2020, 2022).[3] We then analyze the impact of this measure on startup performance outcomes – differentiating between the amount of funds raised, acquisition and IPO. In estimating this relationship between knowledge-base difference and performance outcomes we take a step-wise

[2] *https://www.dimensions.ai/*

[3] We cross-validate our measure with others, which we describe in more detail in the data section. There is strong correlation with these alternative proxies.

approach. We first control for as many confounding factors as possible including a wider set of fixed effects, and then perform several subsequent tests to show that our results remain robust to different empirical and sample specifications. In addition, we provide evidence that selection is unlikely the main driver of our results, and implement an instrumental variable approach. The results remain consistent throughout with our more naive model.

Our analyses reveal noteworthy patterns. Perhaps contrary to what we may expect, we find that startups that are closest to their founder's academic knowledge base are less likely to be acquired and less likely to raise a significant amount of funds: a 10p.p increase in founders' use of their previous academic work when creating their startups is associated with a 2.7% lower probability of acquisition and a 4.1% lower probability of raising more than \$10 million (75^{th} percentile) in the first five years since inception. We do not find any differential effect on the likelihood to IPO. We show that these results do not appear to be driven by lower invention quality nor more nascent technology. Instead, the mechanism behind these findings seems to lie in the *nature* of the scientific knowledge utilized. Our analysis of the papers cited by startups' patents reveals that startups relying more on their founders' academic work tend to harness narrower and more specialized knowledge. While indicative of deep expertise, this may pose challenges for incumbents and venture capitalists (VCs) alike. For incumbents, the challenge arises from the difficulty of integrating highly specialized technology with any complementary assets they may possess. The narrower and more specialized nature of the knowledge makes the technology less adaptable or compatible with existing systems and strategies, reducing its appeal as an acquisition target (Arora, Belenzon and Suh (2022), Polidoro Jr and Yang (2021)), which is consistent with the lower acquisition likelihood we observe. For VCs, the challenge may stem from the intricacy of evaluating the technology's potential for scalability and market impact, in line with the lower probability of raising a significant amount of funds. At the same time, this mechanism is also consistent with the non-significant estimate we find on IPO. Indeed, once a firm has reached this stage, most uncertainty around the technology has been resolved and there is less need to integrate the technology with complementary assets of potential acquirers since the goal is to be a standalone company. In general, IPOs are more focused on the market potential and growth prospects of the company as a whole. Overall, these results highlight that the success of academic startups is influenced not only by the incorporation of scientific knowledge, but also by the specific type and depth of the knowledge that they use. In particular, our findings stress the importance of considering the *nature* of scientific knowledge (i.e., how specialized to the founder) when founding and developing startups as a strategic choice. Beyond generating novel ideas, considerations of integration challenges cannot be ignored, especially provided that the dominant exit strategy of bio-medicine startups is acquisition (Aggarwal and Hsu (2014)). As such, our results carry critical strategic and practical implications for these ventures.

In addition, provided that academic entrepreneurship is a departure from the more traditional academic job description (Cohen et al., 2020), we further examine potential ramifications for academics' research output. Ideally, commercial and academic activities should be reinforcing, such as work focusing on understanding the impact of academic patenting (Azoulay et al., 2009; Crespi et al., 2011) or university licensing (Thursby and Thursby, 2011) on university research suggests. Though similar in the sense that these are all commercial activities, entrepreneurship is likely a stronger departure from academia, so it is not guaranteed the relationship is similar. In fact, the much scarcer literature on academic spinouts and its relationship with research output is rather ambiguous. On the one hand, Toole and Czarnitzki (2010) provide evidence of a significant decrease in research performance within their sample of academic entrepreneurs after they begin working in for-profit firms. On the other hand, results from Fini et al. (2022) suggest that entrepreneurship may lead to more impactful research because founders will engage in greater exploration. As such, it remains an empirical question to understand how the extent to which an academic founder exploits their own unique technological knowledge in creating their startup relates to their academic output.

Implementing a staggered difference-in-difference design with individual-level fixed effects, our results indicate that, on average, there is no significant difference in terms of academic output following startup creation between founders who build and those who do not build on their academic work in their startups. However, this masks important heterogeneity. In particular, we find a decrease in the number of publications and top publications after startup creation for founders whose firm's knowledge base is closest to their academic work compared to founders who are more distant, highlighting the potential trade-off generated by the dual involvement in academia and industry.

Several mechanisms could be at play. A possible one could be that these results are capturing an escalation of commitments (Arkes and Blumer, 1985; Staw, 1976) leading to negative performance outcomes in both entrepreneurship and academia for founders relying the most on their own academic work. Alternatively, these founders could strategically shift their focus towards Intellectual Property (IP), leading them to prioritize the protection and commercialization of their specialized knowledge and to keep greater control over what they publicly disclose. Additional findings suggest that the latter mechanism might be at play: we find that founders who rely more on their previous academic work experience a) a decrease in the number of co-authors after entering entrepreneurship, b) a narrowing of their research breadth but not at the expense of exploration, and c) an increase in their propensity to patent. Combined together, these results provide suggestive evidence of strategic considerations. In particular, academic founders relying more on their own work appear to streamline their research efforts toward areas that have direct applicability to their startup's technology and market strategy, resulting in a focused research agenda that contributes to IP generation. At the same time, in honing in intellectual property, founders may limit their involvement in collaborative projects, either to safeguard and maximize the value of their inventions and/or as a consequence of working on a different type of work. Such refocusing (more narrow, less collaborative, less open) of the research agenda - similar to the patterns we observe in their commercial ventures - could entail broader consequences for the overall production of science and innovation (Aghion et al. (2008)), echoing research showing that innovation has become less disruptive over time (Park et al. (2023))[4].

The observed pattern of founders who rely more on their academic work being less likely to be acquired, raising fewer funds, and exhibiting specific changes in their research behavior after entering entrepreneurship presents a nuanced interplay of strategic decisions. The simultaneous decrease in research output, narrowing of research breadth, and reduced collaboration might be a consequence of a deliberate shift in focus towards protecting and commercializing intellectual property, as evidenced by increased patenting activity. However, this approach may inadvertently contribute to lower acquisition likelihood and fundraising success. While the increase in patenting aligns with a strategy of securing and enhancing the value of innovations, it appears to coincide with a more closed and specialized research approach, impacting the startup's potential for market integration and attractiveness to acquirers and investors (Roche et al. (2020)). This complex interplay underscores the multifaceted nature of strategic decisions

[4]Interviews with academic entrepreneurs reveal that our results are not because academic entrepreneurs want to remain in control of their companies and do not want to go public or be acquired.

made by academic founders simultaneously involved in two distinct realms - entrepreneurship and academia.

2 Conceptual framework

2.1 The role of science for innovation

Scientific knowledge emerges as a powerful catalyst for innovation. Patents that rely on scientific content have more follow-on citations (Sorenson and Fleming (2004)), are more likely to be renewed (Ahmadpoor and Jones (2017)), generate more value for firms (Krieger et al. (2022)) and are more likely to be traded (Arora, Belenzon and Suh (2022)).

The intersection of scientific knowledge and the establishment of startups becomes particularly salient in the context of academic entrepreneurship. The emergence of startups rooted in scientific discoveries reflects the tangible application of academic research within the entrepreneurial landscape. Notably, the prevalence of academic startups has experienced significant growth in the US over the past decades (Rothaermel et al., 2007; Audretsch, 2014; Roche et al., 2020), positioning them as major contributors to innovation, especially in knowledge-intensive sectors (Acs and Audretsch, 1990). Their prominence is particularly evident in fields marked by a strong integration of scientific research and commercialization opportunities (Stokes, 2011) and where appropriability regimes are strong, such as the biological and life sciences.

However, despite the large number of studies on the positive impact of scientific knowledge on innovation, the existing literature has yet to delve deeper into the nuanced role played by the nature of the science employed. While it is widely acknowledged that science is a potential driver of innovation, the specific type and characteristics of the scientific knowledge employed remain largely unexplored. This gap in the literature prompts a critical examination of not only the incorporation of scientific knowledge but also the distinct characteristics and implications associated with the specific *type* of science employed in entrepreneurial ventures.

2.2 A knowledge-based advantage

The central role of resources in shaping not only established firms', but also new ventures' strategic outcomes has been well-documented (Barney, 1991). Entrepreneurial opportunities are presumed to exist primarily because of differences in beliefs of the relative value of resources, where entrepreneurs can leverage their specialist knowledge to create rents (Schumpeter, 1912; Shane and Venkataraman, 2000). In particular, new firms are created by allocating resources to novel ends (Alvarez and Busenitz, 2001), where initial resource endowment, such as social capital, can serve as a critical foundation for the long-term performance of new ventures (Shane and Stuart, 2002). Although a notable body of work (Agarwal et al., 2004; Åstebro et al., 2011; Sørensen, 2007; Stenard and Sauermann, 2016) has made fundamental strides in understanding the impact of capability differentials in shaping both the decision to become an entrepreneur and the entry mode into entrepreneurship, what remains understudied is how initial *technical knowledge endowments* may shape the subsequent performance of firms. Highly unique technical knowledge endowments may present an important factor in establishing a competitive advantage early on in the life of a firm.

Academic entrepreneurship provides a suitable setting to examine the extent to which startups that exploit their founders' technologically unique knowledge perform better. Given the specialized, often tacit knowledge, that academic entrepreneurs possess (Bercovitz and Feldman, 2006) granting them the ability to "recognize the value of new, external information, assimilate it, and apply it to commercial ends [...]" (Cohen and Levinthal 1990, p.128), specialized knowledge is particularly important in the formation of academic startups. As such, they play a critical role in bridging the gap between the academic and the private sector: startups emanating from academia are typically based on scientific advances made within a laboratory, and not surprisingly, target inventions that the private-sector would have not otherwise pursued because of a lack of technical knowledge. From this, we may expect academic startups that are founded based primarily on the scientific work of their founder(s) to fair best provided their advantage based on knowledge about a specific technology. Such knowledge could help place these firms at the frontier of knowledge, may provide a critical source for differentiation, and may thus create more value for customers. Indeed, academic scientists often have extensive knowledge in a narrow field of science no one else possesses (Jones, 2009). From this, it appears, we should expect that *the more a startup exploits the knowledge-base of its founders, the better startup performance.*

2.3 Dual commitment in entrepreneurship and academia

Given their dual involvement in academia and entrepreneurship, academic founders share their time between two core tasks: i) knowledge production (and training) at universities, and ii) creating their venture (Roche, 2023). The extent to which a founder builds a startup on their own unique technological knowledge may thereby also have important consequences for their academic performance. The line of work examining dual involvement of academics in academia and another activity has primarily focused on patenting or licensing (e.g., Azoulay et al. 2009; Crespi et al. 2011; Thursby and Thursby 2011), with the aim of getting a better understanding of the reasons behind professors' involvement in commercial activity (Perkmann et al., 2013), the characteristics of those professors who do (Agrawal and Henderson, 2002), and what the implications of commercialization are for professors' time, knowledge, norms, and resources (Shibayama et al., 2012). The empirical evidence suggests that the most productive academic life scientists are those involved in commercialization (Agrawal and Henderson, 2002) where specifically in the case of biotechnology, an influential stream of research points to the fundamental role "star" scientists play in transferring new academic knowledge to industry (Higgins et al., 2011; Toole and Czarnitzki, 2009; Zucker et al., 2002, 1998). From this we may expect that both activities could be reinforcing.

Other work has investigated the impact of the adoption of commercial attitudes and behaviors by academic researchers on a number of outcomes (Dasgupta and David, 1994; Etzkowitz, 2003; Powell and Owen-Smith, 1998; Powell and Snellman, 2004; Stephan, 2012; Stuart and Ding, 2006), such as sharing behaviors (Shibayama et al., 2012), and shifts in the amount, direction, and quality of scientific research. Some studies highlight potential risks for academic research, such as changes in the content of scientific research toward more applied topics (Blumenthal et al., 1986), a slowing-down of open knowledge diffusion (Murray and Stern, 2007; Nelson, 2004), or even an exodus of academic scientists to industry (Azoulay et al., 2009). Other work finds that commercialization may enhance traditional scholarship (Goldfarb et al., 2009) and does not seem to distract from academic knowledge production (Abramo et al., 2012; Thursby and Thursby, 2007).

The literature on academic spinouts and its relationship with research output remains scarce. On the one hand, work by Czarnitzki and Toole (2010) and Toole and Czarnitzki

(2010) find a significant decrease in research performance among their sample of academic entrepreneurs after they begin working in for-profit firms. On the other hand, Fini et al. (2022) finds that entrepreneurship leads to more impactful research because founders will engage in greater exploration, and Ambos et al. (2008) provide evidence that these successful "ambidextrous" individuals are more highly cited.

Provided the state of extant literature, it is *unclear ex-ante if the extent to which an academic entrepreneur exploits their own unique technological knowledge impacts their scientific productivity and/or research agenda*. Whether the activities are mutually reinforcing or lead to changes in the rate and direction of research remains an empirical question, which we examine in the following sections.

3 Data

3.1 Academic startups dataset

Our analysis relies on a panel dataset of 510 academic startups. To build these data, we closely follow Roche et al. (2020). The dataset was constructed from the population of US startups listed on Crunchbase, which provides significant information about startups' founding team, sectors and financing. Importantly for our analysis, Crunchbase has a broader coverage of technology startups than other sources since it also provides information on startups seeking to raise capital, regardless of whether they have successfully raised the funds, limiting potential selection and survivor bias. We then keep startups in biomedicine (i.e., biotechnology and medical devices) because this sector is tightly linked to academic research and has a relatively high propensity to patent, which enables us to capture the initial knowledge base of a startup. We focus on startups started after 2004 (because Crunchbase has been found to be more accurate in recent years) and before 2015 (in order to have sufficient time to observe outcomes). Information about the founding team was derived from each startup's website, LinkedIn and Bloomberg through extensive manual searches. Among this sample, we retain the 510 startups with at least one professor in the founding team, which we define as academic startups. We match each academic founder to their publication and patenting output using Dimensions AI based on last name, middle name, first name and institution (see Appendix A for more details about the matching algorithm). We uniquely match 597 professors out of the 676 (88%) . Our main specification will further restrict the analysis to startups with at least one patent and for which we can observe pre-entrepreneurship academic output about at least one academic founder. This represents 308 academic startups which form our final sample. We complement this dataset with PatentsView to retrieve patent-level information.

3.2 Calculating the distance between the knowledge-base of a startup and that of its founders

Our main independent variable corresponds to the distance between the knowledge-base of a startup and that of its founders. We conceptualize this measure as a way to capture the extent to which founders rely more on their specialized knowledge when founding their startup. Proximity between the two knowledge-bases should indicate the degree to which founders rely more on their specialized knowledge when founding their startup, while a larger distance should capture a more broad and general approach to founding.

To operationalize our independent variable, we first need to define the knowledge-base of a startup at the time of creation and then compare it to its founders' academic work. We capture the knowledge-base of a startup at the time of creation by considering its first granted patents. We capture founders' academic work by considering the pool of papers they have published before entering entrepreneurship. We then calculate the distance between founders' academic work and the knowledge base of their startup by leveraging the patenting process: when a company applies for a patent, it has to list all the knowledge on which it builds on, including scientific papers. This allows us to differentiate between citations that patents make to the founders' academic work vs citations that patents make to other researchers' work. We then calculate the distance between the knowledge base of a startup and its founders' academic work by computing the percentage of scientific citations that their patents make to their founders' academic papers.

In practice, we take the first granted patent(s) a startup applied to[5] and we match them to the Reliance on Science (RoS) dataset (Marx and Fuegi, 2020, 2022), which provides a publicly-available set of citations from U.S. patents to scientific articles. For each academic article cited by a patent, we create a self-citation dummy equal to 1 when at least one author

[5]For each startup, we consider the granted patent with the earliest application year. In case there are several granted patents with the same earliest application year, we consider all of them.

of the academic article is matched to an inventor with a confidence score above 50.[6] This identifies instances where we can reasonably be confident that at least one inventor of the patent is citing their own academic work. Results are robust to using a more demanding definition where the average confidence score of all authors in matching with an inventor is used. We then calculate the percentage of self-cites at the patent level by dividing the total number of self-cites by the total number of scientific citations.

$$\text{Percentage self-cites} = \frac{\text{Number of self-cites in RoS}}{\text{Total number of cites in RoS}} \tag{1}$$

Finally, we average this measure at the firm level for startups with multiple first patents. The higher the percentage of self-cites, the more founders relied on their previous academic work when creating their venture and so the closer the startup's and founder's knowledge bases are.[7]

3.3 Summary statistics

Our main independent variable of interest is skewed, with an average value of 8.1% (Figure A1 presents the histogram of *Percentage Self-Cites*). A majority of startups have patents that do not rely on their founders' academic work: the treatment variable takes a value of 0, implying a relative high distance between the knowledge base of the startup and that of its founders' academic work.[8]

<Insert Table 1 here>

[6]The match is performed based on last, middle and first name (Marx and Fuegi, 2020, 2022) Results are robust to the use of more stringent thresholds, such as 75.

[7]In order to provide additional support for our knowledge-base proximity variable, we also construct a similarity measure between founders' publications and their venture's patents. To that end, we use the *pmra* probabilistic topic-based model for content similarity developed by Lin and Wilbur (2007) to calculate the similarity between each founder publication abstract published before startup creation and each patent of his venture. We then aggregate this measure at the venture level, using several measures such as the mean, the median or the max. We find a positive correlation between this similarity measure and our treatment variable that relies on the number of self-cites, providing support that our measure captures an overlap between academic and commercial output in a systematic way, rather than, e.g., narcissism. We reiterate this analysis by comparing venture websites and founders' publication abstracts, using the dataset of Guzman and Li (2023) and find similar results. However, we remain cautious of using one of these similarity measures as treatment because they compare different types of goods (publication vs patent or publication vs website) which use a different vocabulary that is not directly comparable between each other and does not entirely reflect the link to knowledge-bases.

[8]Note that there are two cases where the distance value could be null. First, if a startup cites scientific papers but none of them come from the founders' previous academic work (110 startups). Second, if a startup does not cite any scientific papers, which implies that it also does not cite any previous work from its founders (13 startups). In robustness, we show that our results are robust to conditioning on startups with a positive *Percentage Self-Cites* value.

Table 1 displays summary statistics for our sample of 308 firms. Startups have on average 7 patents, and on average 6 cite scientific literature. There are 1.5 professors per startup, for an average team size of 2.3 people. 16% of firms have at least one female founder and 70% have at least one founder who graduated from a top-tier university.[9] 8% of the startups in our sample are acquired and 7% of them go public through an IPO. The amount of funds raised within 5 years of inception is skewed, with an average of $US 16 millions and a median of $US 0.5 million.

4 Estimation strategy and results

Estimating the causal effect of founders' reliance on their academic work when founding a venture is subject to the classic problem of selection: researchers choose whether to enter entrepreneurship and conditional on entering it, choose whether they predominantly rely on their previous work or not. Our main empirical strategy consists in controlling for as many confounding factors as possible (exit market results) and difference-in-difference analysis (academic market). We also provide evidence that "selection into treatment" does not seem to be correlated with variables that would raise obvious endogeneity concerns. While our results cannot be interpreted as definitely causal, we test their robustness in a series of subsequent analyses, including an instrumental variable strategy, that convey a similar story as our baseline results and echo our discussions with academic founders. In what follows, we begin by presenting the OLS results on startup performance. We then proceed to the analysis of founders' performance on the academic market with difference-in-differences models and individual-level fixed effects specifications. In the following section, we will corroborate our results with robustness tests.

4.1 Determinants of the reliance on one's academic work

In Table 2, we start by presenting results related to the probability of relying more or less on previous academic work when founding the venture. Column (1) estimates the association between the *Percentage self-cites* variable and founders' characteristics. Column (2) to (4) successively add the number of publications published before startup creation, the number of

[9]Top-tier universities are determined following the 2016 Academic Ranking of World Universities ("Shanghai Ranking", accessible at shanghairanking.com)

citations received before startup creation[10] and the total amount of grant funding received before startup creation. Overall, conditional on entering entrepreneurship, there does not seem to be any significant differences between professors who rely more or less on their academic work.

<Insert Table 2 here>

4.2 Startup level outcomes on the exit market

4.2.1 Results

In Table 3, we focus on a set of venture outcomes and their relationship with the reliance of a startup's knowledge base on the academic work of its founders. Our regressions are of the form:

$$Y_i = \beta \text{Percentage self-cites}_i + \gamma X_i + \delta_{\text{State}} + \delta_{\text{Founding year}} + \delta_{\text{Sector}} + \epsilon_i$$

where i indexes startups. X_i includes the following list of controls: the log number of patents and the log number of patents relying on scientific literature which both proxy for startup inventive quality, the log of team size calculated with the number of founders at inception which captures the impact of venture size at founding, an indicator equal to 1 if there is at least one female in the founding team and an indicator equal to 1 if at least one founder graduated from a top-tier university. We further control for the average number of publications generated by founders before startup creation.[11] We also add sector, state, and founding-year fixed effects to control for technology, state and founding-year trends that might be correlated with both the outcomes and our treatment variable. We cluster standard errors at the startup level.

Our dependent variables of interest focus on measures that proxy for the success on the exit market. We focus on three measures: (i) an indicator equal to 1 if the startup is in the 75^{th} percentile of the distribution of funds raised within the first 5 years (i.e., $US 10 million or above), (ii) an indicator equal to 1 if the startup was acquired, and (iii) an indicator equal to 1 if the startup went public via an IPO.

[10] Residualized in order to account for differences in publication year

[11] This avoid the bias that could arise from the correlation between founders' number of publications and our treatment variable (e.g., having more publications could be positively correlated with the propensity to cite one's own work) and the correlation between founders' number of publication and success on the entrepreneurship market. In practice, results are not significantly impacted by the inclusion of this variable.

<Insert Table 3 here>

Column (1) shows that a lower distance between a startup's knowledge-base and its founders' academic work (i.e., an increased percentage of self-cites) decreases the probability of being in the top 25% of firms that raise the most amount of funds: a 10p.p decrease in distance (i.e., a 10p.p increase in self-cites) is associated with a 4.1% decrease in the probability of having raised more than $10million within 5 years of inception. Column (2) shows that a lower distance between a startup's knowledge-base and its founders' academic work decreases the likelihood of being acquired: a 10p.p decrease in distance (i.e., a 10p.p increase in self-cites) is associated with a 2.7% decrease in acquisition likelihood. We do not find any significant effect for the probability of IPO.

In the following subsections, we provide suggestive evidence that the exit outcomes are consistent with the fact that founders who rely more on their own academic work tend to create startups whose knowledge-base relies on narrower and more specialized scientific knowledge. Although using science in invention may be an advantage on the market for technologies (Arora, Belenzon and Suh, 2022), the *type* of science being used might play an important role. In particular, the more the startup is based on the narrow and specialized knowledge of their founders, the more likely it may be that potential acquirers do not possess suitable complementary assets rendering the startup less valuable on the exit market (Arora, Fosfuri and Roende, 2022). This makes their technology potentially less easily integrated by the private sector (Bikard et al., 2019). At the same time, this specialized knowledge may not provide any competitive disadvantage on the IPO market where the end goal may not be for the technology to be integrated with another one.

For further robustness, we also implement an instrumental variable approach, where we use the academic network size of founders as an instrument for the *Percentage of self-cites* variable. A bigger academic network could make a founder more likely to rely on others' work vs their own, as they have been exposed to a more diverse and likely broader set of ideas. To proxy for network size, we use the number of unique collaborators affiliated with an academic institution before startup creation[12]. For academic network size to be a suitable instrument, the exclusion restriction cannot be violated. In other words, our instrument should not directly impact the likelihood of acquisition other than through its influence on the extent to which founders rely

[12] When a startup has multiple academic founders, we use the maximum of the values

on their own work when entering entrepreneurship. One issue could be that professors with a bigger academic network size have more prestige and reputation, which correlates with our outcomes. In order to mitigate this threat, we control for the number of publications before entering entrepreneurship and include an indicator equal to 1 if at least one of the professors in the founding team is a full professor.

In addition, to qualify as a suitable instrument, academic network size should not be correlated with a variable that also influences the likelihood of acquisition. One worry could be that individuals with a bigger academic network size also have more ties to the industrial sector. To mitigate this threat, we include an indicator equal to 1 if at least one founder has written an article with someone affiliated with a firm prior to entering entrepreneurship. We also include an indicator equal to 1 if at least one of the founders has already patented prior to entering entrepreneurship. Academic network size could also be correlated with factors such as location, sector, gender or type of university, which could also affect our outcomes. Provided these potential influences, we include controls for these factors: state, and sector fixed effects, gender, and university rank.

Table 4 presents the 2SLS result and shows suggestive evidence that the previous negative correlations between *Percentage self-cites* and the funding and acquisition outcomes are robust to using this network instrument. We do not find significant effect on *IPO*, similar to our OLS result. Our F-statistic is above the rule of thumb with a value of 28.2 and we find that a bigger academic network size is associated with a lower *Percentage self-cites*, in line with the direction we would expect[13]. Note that the coefficients on the reduced form are about 10 times larger than the coefficients from our more naive approach described above which could be due to heterogeneity in the sample we are analyzing. The compliers that we are shifting - those founders who relied less on their own work because they happened to have a bigger academic network - may have higher returns than the average founder of our sample. While we do not aim to use this instrumental variable approach as a way to pinpoint a specific magnitude, we believe it is reassuring to find that a lower distance between a founder's academic work and their startup leads to similar firm performance as our previous OLS result.[14]

[13]The coefficient on the instrument equals $4.85.10^{-5}$

[14]In results left unreported we apply further instruments for the percentage of self-cites (our treatment variable), such as using the number of examiner-added scientific citations. Similar to the judge leniency instrument (Arnold et al., 2018; Dobbie et al., 2018), some examiners have a higher propensity than others to add scientific references. Conditioning on art-unit and application year, the random assignment of examiners to patents can in theory allow us to create an instrument capturing the propensity of examiners to add

<Insert Table 4 here>

4.2.2 Mechanisms

In what follows, we argue that the fact that startups whose knowledge base is closer to that of their founders are less likely to be acquired and to raise a significant amount of funds is related to the *type* of science they build on. We begin by ruling out the possibility that these startups have a technology which is of lower quality, or at a more nascent stage of development. We then analyze the papers on which startups build on in order to characterize the *type* of knowledge they utilize. To that end, we use the DOI and PMID numbers associated with each paper cited by the first patents of a startup and retrieve information about them using Dimensions AI.

Startup level invention outcomes: In Table 5, we explore the relationship between our main independent variable of interest and invention outcomes. We find no association between the number of patents generated by startups and the distance between their knowledge base and that of their founders. Results are similar when looking at the number of patents relying on science.

<Insert Table 5 here>

Table 6 shows that there is no significant difference in the number of forward citations received by patents associated with startups that are more or less distant to their founders' previous academic work. Columns (1) and (2) report results for the first patents associated with the startups, while columns (3) and (4) consider all patents. Overall, these results suggest that the lower likelihood of acquisition experienced by startups whose knowledge base is closer to their founders' academic work is not related to a lower inventive output or a worse innovation quality.

<Insert Table 6 here>

scientific citations. This would instrument for the number of scientific citations associated with each patent (the denominator of our treatment variable) and hence instrument for our treatment variable (*Percentage of self-cites*). However, despite results that show similar magnitudes and signs than our baseline estimates, our first-stage is too weak, probably because examiners add very few citations on average. Ideally, we would have liked to implement the concept of idea twins (Bikard, 2020) where founders with different academic knowledge bases would create startups embodying almost identical ideas, but is not a viable option given our already small sample size.

Stages of technological development: While we do not find any significant difference in inventive quality, the lower success rate experienced by startups whose knowledge base is closer to their founders' academic work could be explained by a more nascent and less developed technology. In order to test this explanation, we construct a proxy for the development time of each startup's technology and examine its relationship with the percentage of self-cites variable. For each startup, we take the application year of its first patent, which can be conceptualized as the time when the technology was ready to be commercialized. We then consider the publication year of the first paper cited by this patent, which can be thought of as the closest approximation to the scientific start date of research around the technology. This should capture when the idea around the technology was first studied in an academic lab but was not yet ready to be commercialized. We then subtract this publication year from the patent application year and use this difference as a proxy for the time needed to develop the technology. The higher this difference, the more time should have been needed to bring a scientific invention to a commercializable state. Results presented in Table A1 show that there is no significant association between the percentage of self-cites variable and the time of technology development. If anything, the magnitude appears negative, implying that startups closer to their founders' academic work are able to transfer their technology faster to the market. Using startup creation year rather than patent application year leads to the same conclusion. In unreported regressions, we classified each papers cited by a patent as being either "basic" or "applied" using the Research Activity Codes provided by the Health Research Classification System (HRCS)[15] and averaged this measure at the startup level. We also do not find any relationship significant on conventional levels.

Characterizing the scientific knowledge startups build on: We now characterize the *type* of science on which startups build on. We focus on two main measures that capture (i) the breadth of the papers startups build on and (ii) the extent to which they embody specialized knowledge.

In order to proxy for the breadth of research, we make use of the *concepts* feature of Dimensions AI. For each paper, Dimensions AI uses a machine learning algorithm that assigns concepts to each paper as well as a relevance score. For each paper, we count the number

[15]See `https://hrcsonline.net/`. The HRCS uses a coding system to classify the full spectrum of biomedical and health research from basic to applied – across all areas of health and disease.

of concepts with a score of 0.5 or above. We then average this measure at the startup level. Results are presented in column (1) of Table 7 and show a significantly negative coefficient, implying that startups relying more on their founders' academic work build on scientific work that is narrower.

For each concept, we also calculated the number of times it was used by other papers and averaged this measure at the startup level. We conceptualize this measure as a proxy for the degree of specialization of the knowledge embodied in a paper: the more specialized the knowledge, the less likely it is to be used in other papers. Results are presented in column (2) of Table 7 and show also a significantly negative coefficient, implying that startups relying more on their founders' academic work build on scientific work that is more specialized. In an unreported regression, we find no significant impact on the number of citations received by papers, implying that this lower likelihood of being used by others is not due to a lower quality of the paper. Column (3) focuses on the citations received by these papers by patents (excluding patents from the focal startup). Interestingly, startups closer to their founders' academic work also seem to build on scientific knowledge that tends to be less used by other patents. This is consistent with the fact that their narrower and more specialized knowledge is less easy to integrate into technological solutions.

<Insert Table 7 here>

Discussion of the results on firm performance Our results regarding firm performance provide evidence that startups whose knowledge base rely more on the previous academic work of their founders are less likely to be acquired and to raise a significant amount of funds, but have no differential impact of going through an IPO. We showed that this result does not seem to be linked to a lower innovation quality, as proxied by the number of patents and the citations these patents receive. We also find no support for the idea that startups closer to their founders' academic work would have built on more nascent technology, as proxied by the time between the publication year of the scientific knowledge used in patents and these patents' application year. These findings suggest that the observed challenges in acquisition and fundraising are not driven by diminished invention or the early-stage nature of the technologies developed by startups that are closer to their founders' academic work. Instead, our attention turned to the *nature* of the scientific knowledge employed. In particular, we find that these startups that rely more on their founders' previous academic work tend to utilize knowledge that is narrower and more specialized, despite being of apparently the same scientific quality. This mechanism is consistent with the lower likelihood of acquisition that they experience, as the integration of technology with complementary assets becomes more complicated when the knowledge base is highly specialized. Similarly, it aligns with the lower propensity to raise a significant amount of funds since VCs may face hurdles in evaluating the potential of highly specialized technologies. This mechanism is also consistent with the absence of detectable effect on IPO, since this outcome is more dependent on overall market potential and growth prospects, where the need for integration with complementary assets is less critical. Our findings underscore a crucial strategic balance for academic founders: leveraging specialized knowledge while considering its market applicability.

4.3 Founder level academic outcomes

4.3.1 Results

Besides their entrepreneurial endeavors, academic entrepreneurs have another job to fulfill: produce public knowledge. In order to assess the relationship between the distance in knowledge bases and academic output, we implement a two-way fixed effects specification:

$$Y_{it} = Treated_i \times Post_t + \alpha_i + \delta_t + X_{it} + \epsilon_{it}$$

where i indexes individuals and t indexes year. In our specification, $Treated_i$ equal 1 if individual i is part of our treatment group (that we define below) and 0 otherwise, while $Post_t$ equal 1 for years following startup creation. X_{it} comprises bins of experience dummies where each dummy has a size of 4 years.[16] We account for differences between the treatment and control groups by including a full set of individual fixed effects. Hence, our identification strategy does not require that professors in the treatment and control groups were similar when they created their startup. Rather, we assume that academic activity in these two groups would have evolved in parallel. We also present evidence in support of this assumption with a staggered event study specification where we interact the treatment group indicator with a

[16] One dummy for years of experience 1-5, one dummy for years of experience 6-10, one dummy for years of experience 11-15, one dummy for years of experience 16-20, one dummy for years of experience 21-25 and one dummy for years of experience above 25. Year 1 of experience corresponds to the first year for which we observe an academic publication.

full set of time leads and lags:

$$Y_{it} = \sum_{\tau \neq -1} \beta_\tau \mathbf{1}[\tau = t - t_S^*] + \alpha_i + \delta_t + X_{it} + \epsilon_{it}$$

where t_S^* indexes the year of startup creation. We will use $\{\beta_\tau, \tau < -1\}$ to identify potential pre-trends.

Our main dependent variables of interest are the number of publications, the number of (year- and field-adjusted) citations and top publications (defined as papers which receive more citations than the average number of citations of other papers in the same area of research and publication year)[17]. Conditional on publishing, we also examine the percentage of papers in which founders appear as first author – which often indicates that the individual performed the major share of the work, and the percentage of papers in which founders appear as last author – which often indicates that the individual was the Principal Investigator of the project.[18]

We keep the 375 founders who have founded only one startup in order to have clear and easily identifiable pre- and post-treatment periods which correspond to the periods before and after entry into entrepreneurship. In our baseline specification, our treatment group is composed of individuals whose startup has patents that cite at least one of their academic papers (i.e., *Percentage self-cites*>0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites*$=0$). We will explore the sensitivity of our results to other definitions of the treatment and control groups in subsequent analysis.

<Insert Table 8 here>

Table 8 provides summary statistics for the treatment and control groups. Panel A indicates that there is no significant difference regarding the number of patents associated with founders who are part of the treatment groups, vs founders who are part of the control group. Panel B explores academic output at the time of startup creation. Both groups have, on average, slightly more than 20 years of experience when they began their entrepreneurial activity. There is no significant difference between the treatment and the control groups with regards to the

[17]This measure is provided by Dimensions AI and is available for PubMed publications. The area of research is defined by publications co-cited with the article of interest and is therefore dynamically created.

[18]A robust social norm in the Life Sciences and Engineering assigns last authorship to the most senior faculty running the lab and leading projects, while first authorship is given to younger individuals involved in writing the paper (Azoulay et al. (2009)), and conducting the bulk of the research work

number of publications, top publications, citations and authorship, prior to startup creation. Though our empirical strategy does not require both groups to be comparable in terms of level, it is reassuring to see that there is no systematic difference in academic output between the treatment and control groups.

Figure A2 in the Appendix displays raw graphs of the main dependent variables of interest as a function of time from (negative values) and since (positive values) startup creation. The number of publications increases for both groups until startup creation. After that, it stabilizes and then seems to slightly decrease for the treatment group, while it first increases and then plateaus at about 10 publications per year for the control group. The number of top publications appears to be decreasing for both groups after startup creation, but to a greater extent for the treatment group. As expected, the number of cites is on a decreasing trend as papers published more recently had less time to gather citations. The percentage of papers written as last author varies between 0.35 and 0.55. It is noteworthy that this value decreases for both groups after startup creation, potentially suggesting some change in research strategy. Overall, from the graphs, it seems that the number of publications and top publications of the treatment group is negatively impacted compared to the control group after entering entrepreneurship.

We now go beyond the raw data and explore results in more detail in a regression framework. Table 9 shows the interaction term of the difference-in-difference regression. There is no significant difference across the five main outcomes between the treatment and control groups.

<Insert Table 9 here>

We also run the event study regression for the six outcomes of interest and plot the coefficients in Figure A3. There is no apparent pre-trend, and the coefficient on the interaction term is not significantly different from 0 in the post-period, except for the number of publications and top publications which seem to be significantly negative in the longer term.

Next, we examine the heterogeneity of this result as a function of the degree of reliance of founders on their academic work. First, we keep the same definition for the control group (*Percentage self-cites*= 0) but we vary the treatment group by looking separately at professors' academic output whose distance with their startup falls in different percentiles of the distribution of the *Percentage self-cites* variable. A plot of the coefficient on the interaction term is presented in Figure 1.

<Insert Figure 1 here>

Results indicate that there is no significant difference between the treatment and control groups for individuals whose academic work is relatively far from the knowledge base of their startup (up to the 75^{th} percentile). However, we find a decrease in the number of publications and top publications for individuals who are above the 90^{th} percentile of closeness. This represents a decrease of about 2 publications and 1 top publication from the mean per year for founders that are closest to the knowledge base of their startup. Taking rough estimates from prior work on the grant dollar to paper relationship, and assuming that this amount should proxy for the lower bound of what a paper is worth (investment should be equal or below expected returns) our back-of-the-envelope calculations suggest that this could represent a loss of value generated from publications of 40,000 - 333,000 dollars per year per academic entrepreneur.[19] Naturally, these numbers are crude and do not take into account the gains from pursuing other activities. These gains could, quite possibly also drive down the costs that go into producing a paper.

We repeat this exercise, this time by varying the threshold below/above which we categorize the control and treatment groups. Figure 2 plots the coefficient of the difference-in-difference interaction term.

<Insert Figure 2 here>

Each point corresponds to a specific threshold for defining the treatment and control groups: for instance the point corresponding to p(1) on the x-axis is the interaction term of the difference-in-difference regression when we define the control (treatment) group as being below (above) the 1^{st} percentile of the treatment variable. Results convey a similar story as before: in particular, professors who tend to be very close to their startups in the knowledge space experience a decrease in the number of publications and top publications.

4.3.2 Mechanisms

So far, our findings indicate that professors who rely extensively on their academic work to create their startup experience (i) poorer exit prospects on the entrepreneurial market and (ii)

[19] Based on Boyack and Börner (2003), Druss and Marcus (2005) and Leydesdorff and Wagner (2009), the range lies between 0.6 and 5 published papers per \$100k in funding. Taking (100,000/5)*2 = 40,000 and (100,000/0.6)*2 = 333,000, we obtain our estimated range.

a decrease in their overall number of publications and top publications after startup creation, compared to individuals who do not rely (or who rely less extensively) on their previous academic work. To get a better understanding of what may be driving poorer performance on the exit market and the changes observed in academia, we examine further individual level outcomes. These are: co-authorship, research direction and patenting activity.

Co-authorship: First, we explore founders' co-authorship networks and how they evolve with entrepreneurship. We calculate the number of unique co-authors that each founder has in a given year - conditional on having at least one paper that year - differentiating between co-authors who share the same affiliation and co-authors working at different institutions. Results are presented in Table A2 and show the difference-in-difference coefficient for the yearly number of unique co-authors (column 1), the yearly number of unique co-authors from the same institution (column 2) and the yearly number of unique co-authors from a different institution (column 3).

Overall, we find no significant difference between the treatment and control groups. However, once again, these results mask significant heterogeneity. As before, we rerun the difference-in-difference regressions by changing the definition of the treatment and control groups. Results are presented in Figure 3 Panel (a), where we keep the definition of the control group as individuals with a *Percentage of self-cites* variable equal to 0 but vary the definition of the treatment group and Figure 3 Panel (b), where we vary the threshold defining the treatment and control groups.

<Insert Figure 3 here>

In both cases, we find that founders who rely extensively on their previous academic output (above the 75$^{\text{th}}$ percentile of the *Percentage of self-cites* variable) experience a decrease in the number of co-authors after startup creation compared to the control group. Interestingly, this result seems driven both by a decrease in the number of co-authors coming from the same institution as well as a decrease in the number of co-authors working at a different university. Overall, this suggests that founders who tend to rely more on their own academic work narrow their network of co-authorship after entering entrepreneurship.

Research direction: Next, we explore changes in research focus after startup creation for professors relying more or less on their previous academic work. The Dimensions AI database uses machine learning techniques to derive concepts in papers' abstracts and rank them based on their relevance on a scale from 0 (not relevant) to 1 (very relevant). We first calculate for each paper the number of concepts with a score above 0.5, which informs us about the diversity of main ideas present in a paper. We then average this measure at the professor-year level. Similarly, among concepts with a score above 0.5, we calculate the number of new concepts, defined as concepts that do not appear in any previous papers published by the founder in the 5 preceding years. We then calculate the share of new concepts by dividing the number of new concepts by the overall number of unique concepts and average this measure at the professor-year level.[20] Figure 4 shows the interaction term of the difference-in-difference regression for both the number of unique concepts and the share of new concepts, looking at heterogeneity across founders.

<Insert Figure 4 here>

As before, Figure 4 Panel (a) keeps the control group as being individuals with a *Percentage of self-cites* variable equal to 0 and varies the treatment group definition, while Figure 4 Panel (b) varies the threshold defining the treatment and control groups. In both cases, we find that founders who are very close to their startup's knowledge base tend to decrease the number of concepts they use in their papers after startup creation compared to the treatment group. However, there does not seem to be any difference in the share of new concepts they use in their papers, suggesting that this narrowing does not seem to come at the expense of topic exploration.

We then reiterate the analysis using the number of unique Medical Subject Headings (MeSH)[21] terms and the share of new MeSH terms[22] used by founders yearly. Results are presented in Figure 5 and show similar patterns: founders who are very close to their startup's knowledge base tend to decrease the number of terms they use in their papers after startup

[20] Using the share rather than the number of new concepts allows to control for differences in research breadth. E.g., 5 new concepts could be high if the research area of a paper is narrow with an average of e.g., 10 unique concepts (50% of novelty) but is low if the research area of a paper is broad with e.g., 50 unique concepts (10% of novelty).

[21] MeSH terms are a set of keywords maintained by the National Library of Medicine that indexes the intellectual content of PubMed articles. See for example Azoulay et al. (2019).

[22] New MeSH terms are those that haven't been used in the previous 5 years.

creation compared to the treatment group, but there is no significant impact on the share of new terms used.

<Insert Figure 5 here>

Overall, this suggests that founders who tend to rely more on their own academic work experience a change in their research agenda after founding, focusing on a narrower set of ideas. This interpretation is in line with canonical theoretical models suggesting that private sector research is more focused than academic research, where the fundamental trade-off lies in giving up creative control for higher, more certain payoffs (Aghion et al., 2008). This may occur through specialization around a smaller, defined, sliver of the research "pie" (Jones, 2009). It is interesting to find that this narrowing does not seem to come at the expense of exploration.

Patenting: Following the same difference-in-difference strategy, we also explore changes in patenting activity following startup creation. We use Dimensions AI to match each founder to their patenting output and reiterate the difference-in-difference analysis. Results are presented in Figure 6. Varying the definition of treatment and control groups, we find that founders who rely more extensively on their previous academic output increase their patenting activity after startup creation compared to founders who rely less on it.

<Insert Figure 6 here>

Discussion of the results on the academic market: Overall, we observe significant notable shifts in the behavior of founders who rely more on their previous academic work. Notably, these founders experience a decrease in the number of co-authors after entering entrepreneurship, indicating a shift towards more independent work. Simultaneously, there is a narrowing of their research breadth but not at the expense of exploration, suggesting a strategic focus on more targeted areas while still embracing novel concepts. Furthermore, there is a pronounced increase in their propensity to patent, emphasizing a strategic orientation towards protecting and commercializing their specialized knowledge. While suggestive only, these findings collectively align with what appears to be founders' strategic move to prioritize the protection and commercialization of their innovations, allowing them to maintain greater control over what they publicly disclose. In our interviews, an acquirer referred to this as "grafting on new technology to the [firm's established] tech[nology] stem."

Naturally, this is not necessarily the exclusive driver behind our results and other forces might be at play. In particular, founders who are closer to their startups' knowledge base could also suffer from an escalation of commitments: they experience more difficulties on the entrepreneurship side, which leads them to progressively invest less of their time in academia, shrink their lab size and hence write less papers. This mechanism is however less consistent with the increase in patenting activity and the fact that scientific exploration is not compromised. While we abstain from parsing out the extent of influence exerted by the two potential mechanisms, our interviews suggest that escalating commitment is not the main driver of these results. The following example is particularly illustrative of the conversations we have had. A founder relatively close to his startup's knowledge base explained that he decreased his collaborations with colleagues, from within and outside his department, because the type of research he was conducting had shifted and was more catered towards his new entrepreneurship project. He mentioned this was also the reason why he narrowed his research agenda, though it did not come at the expense of exploration because of the interdisciplinary nature of the project. The fact that patenting activity increases and the share of new knowledge concepts used in papers does not significantly decrease tend to give greater support to the mechanism related to a shift towards Intellectual Property.

5 Robustness

Goodman-Bacon (2021) provide evidence that when there is variation in treatment timing, the two-way fixed effects estimator is a weighted average of all possible 2×2 difference-in-difference estimators, which can lead to bias when there is dynamic treatment effect. Hence, we re-estimate the more traditional difference-in-difference results by creating several datasets (one per year) where we drop already-treated units from the control group, and stack these datasets together. Results are presented in Table A3 and appear robust to this specification (i.e., on average, there is no significant difference in academic output between founders who rely on their academic work and those who do not rely on it). Next, we provide evidence that our results are robust to excluding startups that do not cite any of their founders' academic work, i.e., startups with a *Percentage of self-cites* variable equals to 0. This leaves us with a sample of 185 startups, with a mean *Percentage of self-cites* variable of 13.5% (see Figure A4 in the Appendix for an histogram of the treatment variable). Table A4, A5 and A6 replicate the previous regressions respectively for the exit outcomes, the innovation outcomes, and the characterization of the science used in startups' patents. The results are similar to what we previously found.

We then explore the robustness of our results regarding academic outcomes. As before, we replicate our heterogeneity analysis, but focus on the difference-in-difference specification where the treatment and control groups are defined by the threshold (since we can't define the control as having a treatment variable equal to 0 because these observations are excluded). Results are presented in Figure A5 and are very similar to the main ones: the number of publications and top publications for founders who rely extensively on their previous academic work decreases post-startup creation compared to the control group. Figures A6, A7 and A8 focus on the mechanisms, exploring co-authorship, research breadth and patenting activity. Similar as before, we find a decrease in co-authorship, a narrowing of research breadth and an increase in patenting activity for founders who rely more on their academic work when creating their startup.

We also replicate our results by using all patents pertaining to a startup to calculate the distance measure. Said differently, we calculate the percentage of self-cites at the startup level by considering all its patents, and not just its first ones. This measure captures a more dynamic idea of the distance between founders' academic work and the knowledge base of their startup. Results on the exit market are unchanged. Table A7 shows very similar magnitudes as our baseline results, with a 10p.p increase in founders' reliance on their own academic work being associated with a 60% lower likelihood to raise more than $10 million of funds and a 26% lower likelihood of getting acquired. Again, we do not find any significant difference in innovation quality as proxied by number of patents (Table A8) or the number of citations received by these patents (Table A9). We also do not find any significant difference with regards to technological development. We also consider all the papers cited by all the patents associated with each startup in order to characterize the science they use. Table A10 shows the results of this exercise. We find again that the concepts used in the papers cited by the patents are less used by other researchers and less likely to be cited by other patents. The only difference compared to our baseline results concerns the number of concepts used in papers, for which we lose significance.

6 Discussion and conclusion

Herein, we examine how the extent to which an academic founder exploits their own technologically unique knowledge in the formation of their startup relates to entrepreneurial success and academic output after creation. To do so, we create a dataset comprising 510 academic startups corresponding to 676 founder-professors in biomedicine. We find evidence that startups that rely more on their founders' academic work are less likely to raise a significant amount of funds and to be acquired, while we do not find significant impact on their probability to go through an IPO. This does not seem to result from lower invention quality or an earlier stage of technological development. Rather, we find evidence that startups which rely more on their founders' academic work build on a narrower and more specialized knowledge which might pose challenges for incumbents and investors alike. Our findings highlight that the success of academic startups is influenced not only by the incorporation of scientific knowledge, but also by the specific type and depth of the knowledge that they use, with critical strategic and practical implications for these ventures.

In a next step, we examine the difference in academic output between founders who rely more on less on their academic work when creating their startup. Assessing the consequences of entrepreneurial strategy on founders' core output is critical to better comprehend potential trade-offs associated with commercial engagement. While we do not find any significant difference in academic output after startup creation between these two groups on average, our results exhibit significant heterogeneity. In particular, founders who rely more extensively on their previous academic work show a decrease in the number of publications and top publications after entering entrepreneurship. In subsequent analysis, we find evidence that those founders who rely more on their own work to create their startup tend to decrease their collaborations with other scientists, narrow their research breadth but not at the expense of exploration and increase their patenting activity. While we cannot draw any definitive causal conclusion, we interpret these findings as providing suggestive evidence of strategic considerations. In particular, academic founders relying more on their own work appear to streamline their research efforts toward areas that have direct applicability to their startup's technology and market strategy, resulting in a focused research agenda that contributes to IP generation. Simultaneously, the strategic focus on intellectual property may limit involvement in collaborative projects, whether to safeguard and maximize the value of innovations or as a consequence of working on different types of projects. This refocusing—characterized by a more narrow, less collaborative, and less open research agenda—mirrors patterns observed in their commercial ventures and holds potential broader consequences for the overall production of science. Our extensive interviews with academic founders confirm our conjectures – at least anecdotally.

Combining the results for firm and individual level performances presents a nuanced interplay of strategic decisions among academic founders. The observed pattern of founders who heavily rely on their academic work being less likely to be acquired and raising fewer funds, while exhibiting specific changes in their research behavior after entering entrepreneurship, presents a strategic decision-making in a complex landscape. The simultaneous decrease in research output, narrowing of research breadth, and reduced collaboration may be a consequence of a deliberate shift in focus toward protecting and commercializing intellectual property, evident in increased patenting activity. However, this approach may inadvertently contribute to lower acquisition likelihood and fundraising success. While the increase in patenting aligns with a strategy of securing and enhancing the value of innovations, it coincides with a more closed and specialized research approach, impacting the startup's potential for market integration and attractiveness to acquirers and investors. This multifaceted interplay underscores the intricate nature of strategic decisions made by academic founders, where the pursuit of protection and control may have implications not only for their ventures' success but also for broader entrepreneurial outcomes.

As such, our study may aid in developing more nuanced and targeted policies for academic innovation, acknowledging that some founders may benefit from deeper involvement and support in order to be successful. Moreover, our work directly replies to the call of understanding what the role of intermediaries (e.g., Technology Transfer Offices) may be (`https://twitter.com/heidilwilliams_/status/1646889128242593792`). Perhaps, more counter-intuitively, is that to be successful on the exit market, founders may be better advised to build their nascent firms around a knowledge-base beyond their own expertise, which to some academic scientist represents a departure from their occupational mental model (El-Awad et al., 2022). New programs, such as the Schmidt Science Fellows, which aims at "identifying, developing, and amplifying the next generation of science leaders, building a community of scientists and supporters of interdisciplinary science, and leveraging this network to drive sector-wide change" through postdoctoral fellowships may provide a fruitful approach towards fostering such outcome (`https://schmidtsciencefellows.org/overview/`).

Our study is not without limitations. Although we take great care in controlling for potential confounders and show that our results are robust to different specifications, we cannot formally claim causality. Future research could try to find clever ways to implement quasi-experimental strategies that could help deal with identification challenges. Future work could also try to study features other than knowledge distance and their impact on entrepreneurial and academic success. Given the complexity of firm creation, and persistent consequences of early decisions around the resources used as the foundation of the firm (Geroski et al., 2010), opportunities for follow-on work abound.

References

Abramo, G., D'Angelo, C. A., Ferretti, M. and Parmentola, A. (2012), 'An individual-level assessment of the relationship between spin-off activities and research performance in universities', *R&D Management* **42**(3), 225–242.

Agarwal, R., Echambadi, R., Franco, A. M. and Sarkar, M. B. (2004), 'Knowledge transfer through inheritance: Spin-out generation, development, and survival', *Academy of Management journal* **47**(4), 501–522.

Aggarwal, V. A. and Hsu, D. H. (2014), 'Entrepreneurial exits and innovation', *Management Science* **60**(4), 867–887.

Aghion, P., Dewatripont, M. and Stein, J. C. (2008), 'Academic freedom, private-sector focus, and the process of innovation', *The RAND Journal of Economics* **39**(3), 617–635.

Agrawal, A. and Henderson, R. (2002), 'Putting patents in context: Exploring knowledge transfer from MIT', *Management Science* **48**(1), 44–60.

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

Alvarez, S. A. and Busenitz, L. W. (2001), 'The entrepreneurship of resource-based theory', *Journal of Management* **27**(6), 755–775.

Ambos, T. C., Mäkelä, K., Birkinshaw, J. and d'Este, P. (2008), 'When does university research get commercialized? creating ambidexterity in research institutions', *Journal of Management Studies* **45**(8), 1424–1447.

Arkes, H. R. and Blumer, C. (1985), 'The psychology of sunk cost', *Organizational Behavior and Human Decision Processes* **35**(1), 124–140.

Arnold, D., Dobbie, W. and Yang, C. S. (2018), 'Racial bias in bail decisions', *The Quarterly Journal of Economics* **133**(4), 1885–1932.

Arora, A., Belenzon, S. and Suh, J. (2022), 'Science and the market for technology', *Management Science* **68**(10), 7176–7201.

Arora, A., Fosfuri, A. and Roende, T. (2022), Caught in the middle: The bias against startup innovation with technical and commercial challenges, Working Paper 29654, National Bureau of Economic Research. **URL:** *http://www.nber.org/papers/w29654*

Åstebro, T., Chen, J. and Thompson, P. (2011), 'Stars and misfits: Self-employment and labor market frictions', *Management Science* **57**(11), 1999–2017.

Audretsch, D. B. (2014), 'From the entrepreneurial university to the university for the entrepreneurial society', *The Journal of Technology Transfer* **39**, 313–321.

Azoulay, P., Ding, W. and Stuart, T. (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, P., Fons-Rosen, C. and Zivin, J. S. G. (2019), 'Does science advance one funeral at a time?', *American Economic Review* **109**(8), 2889–2920.

Barney, J. (1991), 'Firm resources and sustained competitive advantage', *Journal of Management* **17**(1), 99–120.

Bercovitz, J. and Feldman, M. (2006), 'Entpreprenerial universities and technology transfer: A conceptual framework for understanding knowledge-based economic development', *The Journal of Technology Transfer* **31**, 175–188.

Bhaskarabhatla, A. and Hegde, D. (2014), 'An organizational perspective on patenting and open innovation', *Organization Science* **25**(6), 1744–1763.

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

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

Blumenthal, D., Epstein, S. and Maxwell, J. (1986), 'Commercializing university research', *New England Journal of Medicine* **314**(25), 1621–1626.

Boyack, K. W. and Börner, K. (2003), 'Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers', *Journal of the American Society for Information Science and Technology* **54**(5), 447–461.

Cohen, W. M. and Levinthal, D. A. (1990), 'Absorptive capacity: A new perspective on learning and innovation', *Administrative Science Quarterly* **35**(1), 128–152.

Cohen, W. M., Sauermann, H. and Stephan, P. (2020), 'Not in the job description: The commercial activities of academic scientists and engineers', *Management Science* **66**(9), 4108–4117.

Colombo, M. G. and Piva, E. (2012), 'Firms' genetic characteristics and competence-enlarging strategies: A comparison between academic and non-academic high-tech start-ups', *Research Policy* **41**(1), 79–92.

Conti, A. and Roche, M. (2021), 'Lowering the bar? external conditions, opportunity costs, and high-tech start-up outcomes', *Organization Science* **32**(4), 965–986.

Crespi, G., D'Este, P., Fontana, R. and Geuna, A. (2011), 'The impact of academic patenting on university research and its transfer', *Research Policy* **40**(1), 55–68.

Czarnitzki, D. and Toole, A. A. (2010), 'Is there a trade-off between academic research and faculty entrepreneurship? evidence from us nih supported biomedical researchers', *Economics of Innovation and New Technology* **19**(5), 505–520.

Dasgupta, P. and David, P. A. (1994), 'Toward a New Economics of Science', *Research Policy* **23**(5), 487–521.

Dobbie, W., Goldin, J. and Yang, C. S. (2018), 'The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges', *American Economic Review* **108**(2), 201–240.

Druss, B. G. and Marcus, S. C. (2005), 'Tracking publication outcomes of national institutes of health grants', *The American journal of medicine* **118**(6), 658–663.

El-Awad, Z., Brattström, A. and Breugst, N. (2022), 'Bridging cognitive scripts in multidisciplinary academic spinoff teams: A process perspective on how academics learn to work with non-academic managers', *Research Policy* **51**(10), 104592.

Etzkowitz, H. (2003), 'Research groups as "quasi-firms": The invention of the entrepreneurial university', *Research Policy* **32**(1), 109–121.

Fini, R., Perkmann, M. and Ross, J.-M. (2022), 'Attention to exploration: The effect of academic entrepreneurship on the production of scientific knowledge', *Organization Science* **33**(2), 688–715.

Geroski, P. A., Mata, J. and Portugal, P. (2010), 'Founding conditions and the survival of new firms', *Strategic Management Journal* **31**(5), 510–529.

Goldfarb, B., Marschke, G. and Smith, A. (2009), 'Scholarship and inventive activity in the university: Complements or substitutes?', *Economics of Innovation and New Technology* **18**(8), 743–756.

Goodman-Bacon, A. (2021), 'Difference-in-differences with variation in treatment timing', *Journal of Econometrics* **225**(2), 254–277.

Grimaldi, R., Kenney, M., Siegel, D. S. and Wright, M. (2011), '30 years after bayh–dole: Reassessing academic entrepreneurship', *Research Policy* **40**(8), 1045–1057.

Guzman, J. and Li, A. (2023), 'Measuring founding strategy', *Management Science* **69**(1), 101–118.

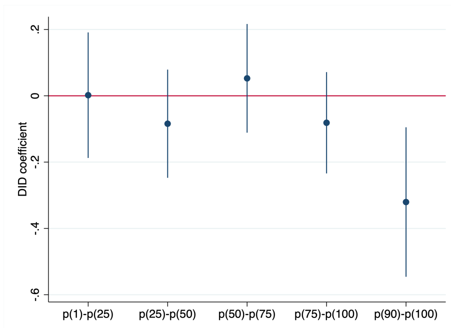
Higgins, M., Stephan, P. and Thursby, J. (2011), 'Conveying Quality and Value in Emerging Industries: Star Scientists and the Role of Signals in Biotechnology', *Research Policy* **40**(4), 605–617.
Hsu, D. H., Roberts, E. B. and Eesley, C. E. (2007), 'Entrepreneurs from technology-based universities: Evidence from MIT', *Research Policy* **36**(5), 768–788.
Jones, B. 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.
Krieger, J. L., Schnitzer, M. and Watzinger, M. (2022), Standing on the shoulders of science, Technical report, Harvard Business School Working Paper.
Leydesdorff, L. and Wagner, C. (2009), 'Research funding and research output: A bibliometric contribution to the us federal research roadmap', *J. of Informetrics* .
Lin, J. and Wilbur, W. J. (2007), 'Pubmed related articles: A probabilistic topic-based model for content similarity', *BMC bioinformatics* **8**(1), 1–14.
Marx, M. and Fuegi, A. (2020), 'Reliance on science: Worldwide front-page patent citations to scientific articles', *Strategic Management Journal* **41**(9), 1572–1594.
Marx, M. and Fuegi, A. (2022), 'Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations', *Journal of Economics & Management Strategy* **31**(2), 369–392.
Mowery, D. C., Nelson, R. R., Sampat, B. and Ziedonis, A. (2004), *Ivory Tower and Industrial Innovation*, Palo Alto: Stanford Business Books.
Murray, F. and Stern, S. (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.
National Science Board (2018), *Science and Engineering Indicators 2016*, Arlington, VA: National Science Foundation.
Nelson, R. R. (2004), 'The market economy, and the scientific commons', *Research Policy* **33**(3), 455–471.
Park, M., Leahey, E. and Funk, R. J. (2023), 'Papers and patents are becoming less disruptive over time', *Nature* **613**(7942), 138–144.
Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A. and Sobrero, M. (2013), 'Academic engagement and commercialisation: A review of the literature on university–industry relations', *Research Policy* **42**(2), 423–442.
Polidoro Jr, F. and Yang, W. (2021), 'Corporate investment relationships and the search for innovations: An examination of startups' search shift toward incumbents', *Organization Science* **32**(4), 909–939.
Powell, W. W. and Owen-Smith, J. (1998), 'Universities and the Market for Intellectual Property in the Life Sciences', *Journal of Policy Analysis and Management* **17**(2), 253–277.
Powell, W. W. and Snellman, K. (2004), 'The knowledge economy', *Annual Review of Sociology* **30**(1), 199–220.
Roach, M. (2017), 'Encouraging entrepreneurship in university labs: Research activities, research outputs, and early doctorate careers', *PLOS ONE* **12**(2).
Roche, M., Conti, A. and Rothaermel, F. T. (2020), 'Different founders, different venture outcomes: A comparative analysis of academic and non-academic startups', *Research Policy* **49**(10), 104062.
Roche, M. P. (2023), 'Academic entrepreneurship: Entrepreneurial advisors and their advisees' outcomes', *Organization Science* **34**(2), 959–986.
Rothaermel, F. T., Agung, S. D. and Jiang, L. (2007), 'University entrepreneurship: a taxonomy of the literature', *Industrial and corporate change* **16**(4), 691–791.
Schumpeter, J. A. (1912), 'Theorie der Wirtschaftlichen Entwicklung: Eine Untersuchung ueber Unternehmergewinn', *Kaptial, Kredit, Zins und den Konjunkturzyklus, Duncker&Humblot, Berlin* .
Shah, S. K. and Pahnke, E. C. (2014), 'Parting the ivory curtain: Understanding how universities support a diverse set of startups', *The Journal of Technology Transfer* **39**, 780–792.
Shane, S. and Stuart, T. (2002), 'Organizational endowments and the performance of university start-ups', *Management Science* **48**(1), 154–170.
Shane, S. and Venkataraman, S. (2000), 'The promise of entrepreneurship as a field of research', *Academy of Management Review* **25**(1), 217–226.
Shibayama, S., Walsh, J. P. and Baba, Y. (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, V. K., Singh, P., Karmakar, M., Leta, J. and Mayr, P. (2021), 'The journal coverage of web of

science, scopus and dimensions: A comparative analysis', *Scientometrics* **126**, 5113–5142.
Sørensen, J. B. (2007), 'Bureaucracy and entrepreneurship: Workplace effects on entrepreneurial entry', *Administrative Science Quarterly* **52**(3), 387–412.
Sorenson, O. and Fleming, L. (2004), 'Science and the diffusion of knowledge', *Research policy* **33**(10), 1615–1634.
Staw, B. M. (1976), 'Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action', *Organizational Behavior and Human Performance* **16**(1), 27–44.
Stenard, B. S. and Sauermann, H. (2016), 'Educational mismatch, work outcomes, and entry into entrepreneurship', *Organization Science* **27**(4), 801–824.
Stephan, P. E. (2012), *How Economics Shapes Science*, Vol. 1, Cambridge: Harvard University Press.
Stokes, D. E. (2011), *Pasteur's quadrant: Basic science and technological innovation*, Brookings Institution Press.
Stuart, T. and Ding, W. (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.
Thursby, J. G. and Thursby, M. C. (2007), 'University Licensing', *Oxford Review of Economic Policy* **23**(4), 620–639.
Thursby, J. G. and Thursby, M. C. (2011), 'Faculty participation in licensing: implications for research', *Research Policy* **40**(1), 20–29.
Toole, A. A. and Czarnitzki, D. (2009), 'Exploring the relationship between scientist human capital and firm performance: The Case of Biomedical Academic Entrepreneurs in the SBIR Program', *Management Science* **55**(1), 101–114.
Toole, A. A. and Czarnitzki, D. (2010), 'Commercializing science: Is there a university "brain drain" from academic entrepreneurship?', *Management Science* **56**(9), 1599–1614.
Zucker, L. G., Darby, M. R. and Armstrong, J. S. (2002), 'Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology', *Management Science* **48**(1), 138–153.
Zucker, L. G., Darby, M. R. and Brewer, M. B. (1998), 'Intellectual human capital and the birth of U.S. biotechnology enterprises', *The American Economic Review* **88**(1), 290–306.

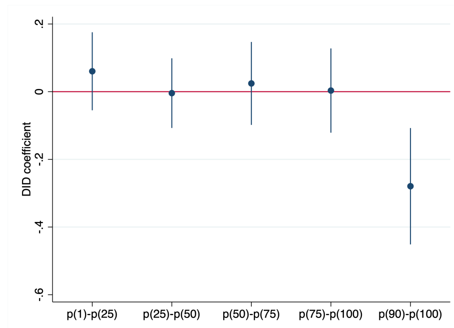
Figures

Figure 1: Heterogeneity Analysis of Academic Output, varying the definition of the treatment group

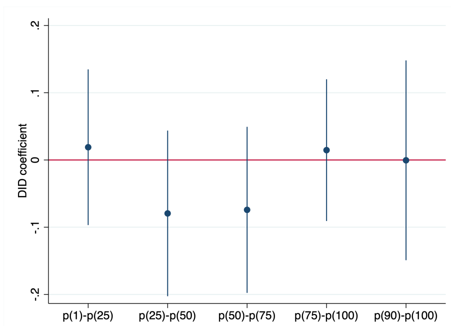
(a) Log(1+Publications)



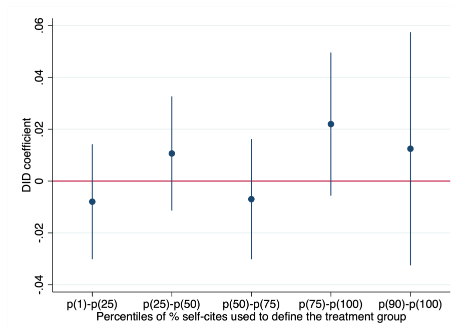
(b) Log(1+Top Publications)



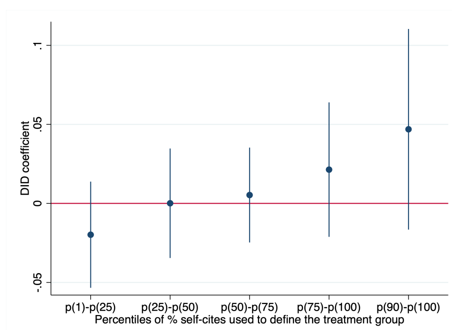
(c) Log(1+Citations)



(d) Log(1+% papers as first author)



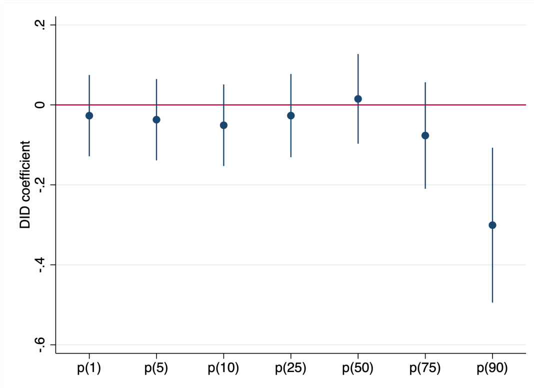
(e) Log(1+% papers as last author)



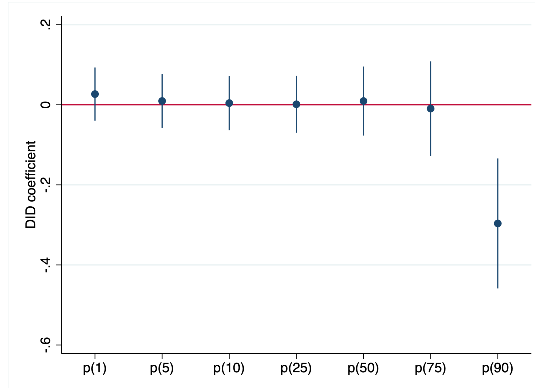
Notes: This figure shows the difference-in-difference coefficients varying the definition of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. For instance, p(1)-p(25) means that all individuals whose *Percentage self-cites* variable falls between the 1^{st} and the 25^{th} percentiles are included in the treatment group. The control group does not vary and includes all individuals with Percentage self-cites= 0. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log average number of citations. In Panel (d), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (e), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure 2: Heterogeneity Analysis of Academic Output, varying the threshold to define treatment and control groups

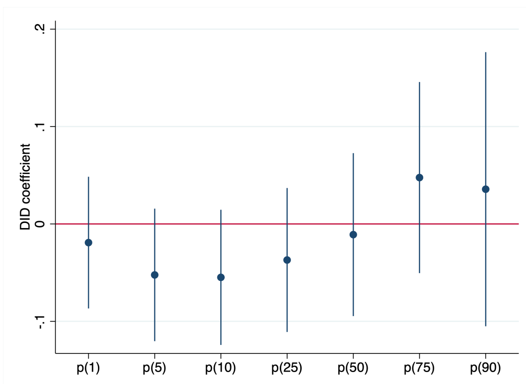
(a) Log(1+Publications)



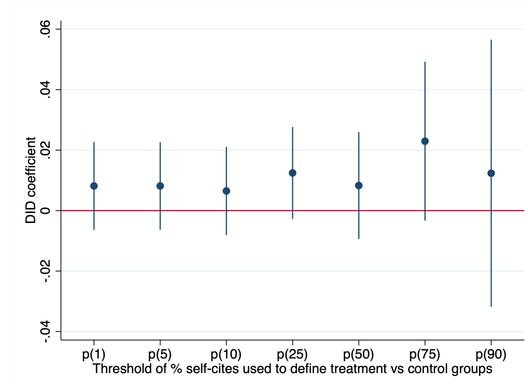
(b) Log(1+Top Publications)



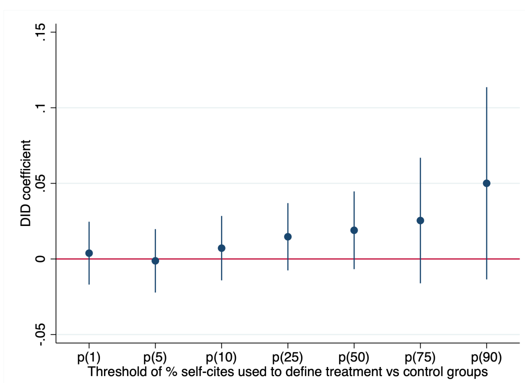
(c) Log(1+Citations)



(d) Log(1+% papers as first author)



(e) Log(1+% papers as last author)

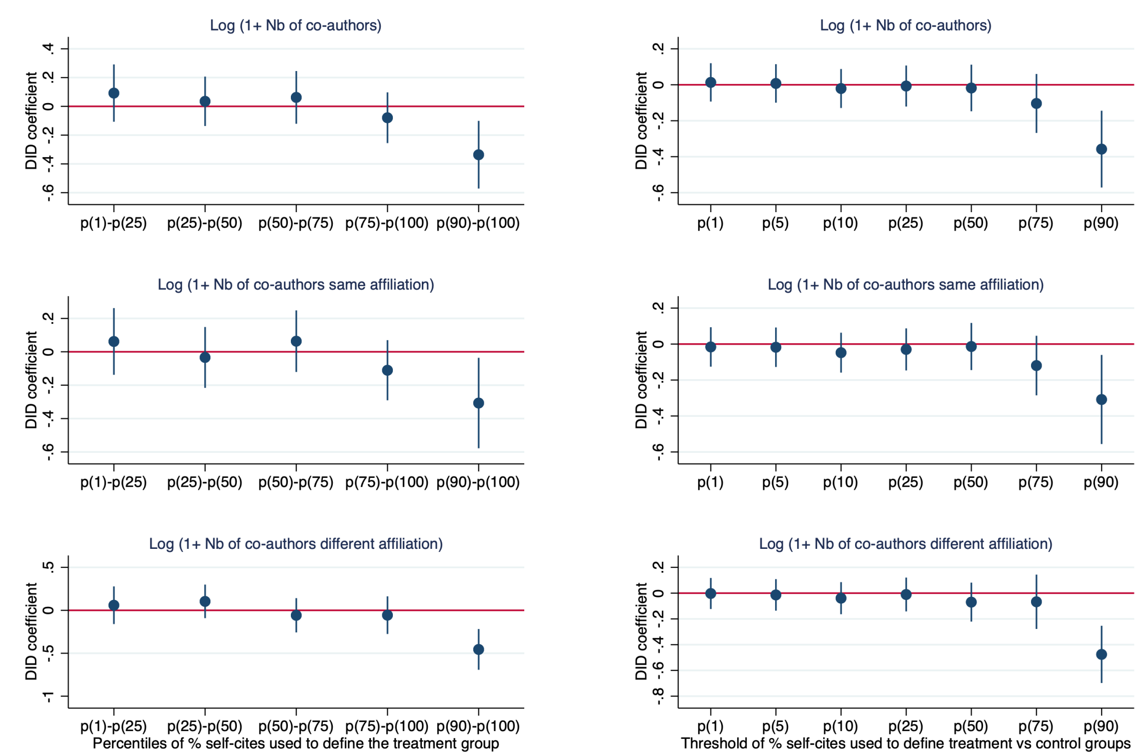


Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log average number of citations. In Panel (d), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (e), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure 3: Heterogeneity Analysis of the Number of Co-Authors

(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups

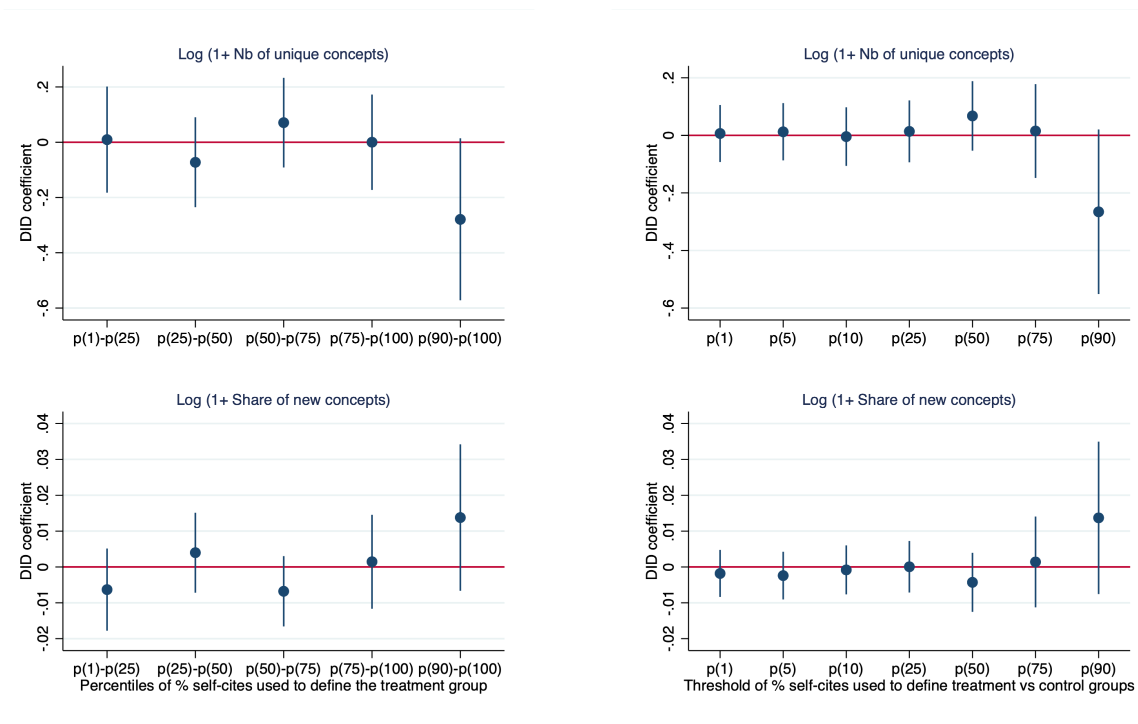


Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique co-authors founders published with in a given year. In the second row, the outcome is the log number of unique co-authors from the same institution founders published with in a given year. In the third row, the outcome is the log number of unique co-authors from a different institution founders published with in a given year. Each bar denotes the 90% confidence interval.

Figure 4: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts

(a) Using the same control group and varying the treatment group

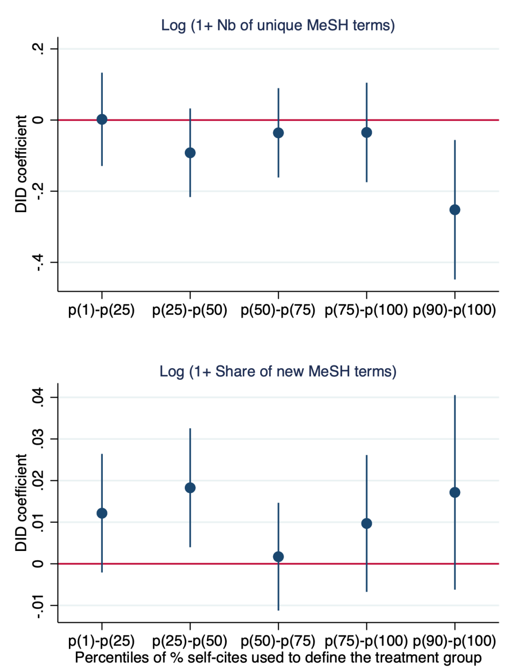
(b) Varying the threshold to define treatment and control groups



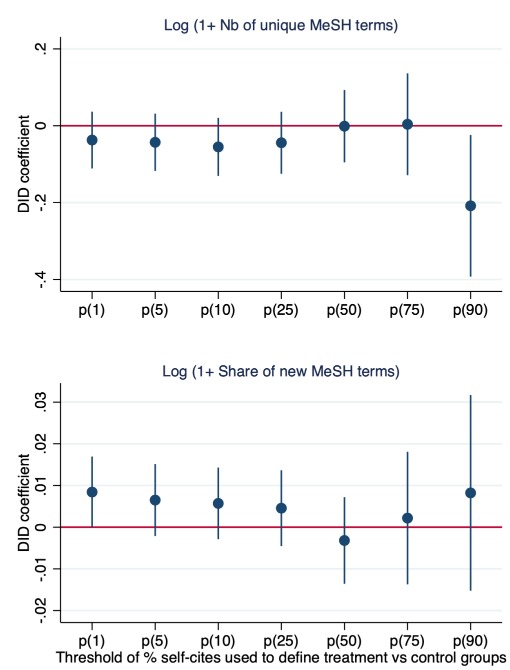
Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* = 0 and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique concepts founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new concepts founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure 5: Heterogeneity Analysis of Research Focus, using MeSH terms

(a) Using the same control group and varying the treatment group



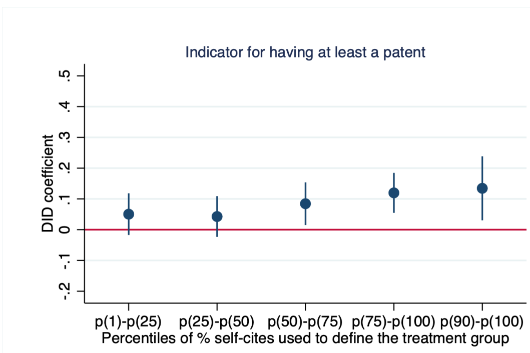
(b) Varying the threshold to define treatment and control groups



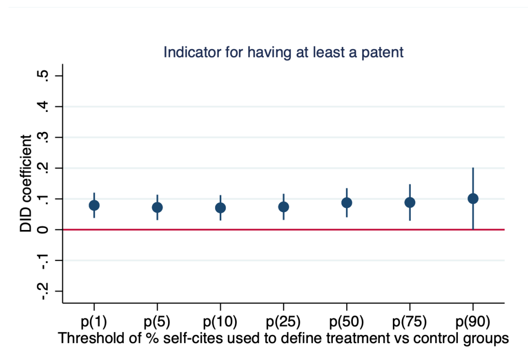
Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, the outcome is the log number of unique MeSH terms founders used in a given year in their papers. In the second row, the outcome is the log number of the share of new MeSH terms founders used in a given year in their papers. Each bar denotes the 90% confidence interval.

Figure 6: Heterogeneity Analysis, Patenting Activity

(a) Using the same control group and varying the treatment group



(b) Varying the threshold to define treatment and control groups



Notes: This figure shows the difference-in-difference coefficients. In panel (a), we keep as control group individuals with *Percentage of self-cites* $= 0$ and vary the definitions of the treatment group. The x-axis shows which observations of the *Percentage self-cites* variable are included in the treatment group. In panel (b), we use different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. The outcome is an indicator variable equal to 1 if the founder filed a patent in a given year and 0 otherwise. Each bar denotes the 90% confidence interval.

Tables

Table 1: Summary statistics, Startup level

	Min	p(50)	Mean	Max
Percentage of self-cites	0	0.02	.08	1
Number patents	1	4	6.7	73
Number patents in RoS	0	3	6.0	67
Number professors	1	1	1.5	6
Team size	1	2	2.3	7
At least one female founder	0	0	0.16	1
At least one top-tier university	0	1	0.70	1
Biotechnology sector	0	1	0.82	1
Founding year	2005	2008	2008	2012
Acquisition	0	0	.08	1
IPO	0	0	.07	1
Amount of funds raised within 5y ($million)	0	.50	15.8	181
Observations	308			

Notes: This table shows summary statistics for our sample of 308 academic startups.

Table 2: Determinants of founders' reliance on previous academic work

	(1)	(2)	(3)	(4)
	Percentage self-cites			
Experience=[6,10]	-0.0178	-0.0179	-0.0179	-0.0175
	(0.0191)	(0.0191)	(0.0192)	(0.0192)
Experience=[11,15]	-0.0209	-0.0210	-0.0213	-0.0193
	(0.0196)	(0.0196)	(0.0196)	(0.0196)
Experience=[16,20]	-0.0103	-0.0104	-0.0101	-0.0104
	(0.0181)	(0.0181)	(0.0182)	(0.0182)
Experience$\geq$20	0.0135	0.0131	0.0148	0.0157
	(0.0184)	(0.0192)	(0.0199)	(0.0198)
Top Institution	0.00256	0.00252	0.00277	0.00237
	(0.0105)	(0.0106)	(0.0106)	(0.0105)
Full prof.	-0.0122	-0.0125	-0.0102	-0.00959
	(0.0134)	(0.0141)	(0.0155)	(0.0154)
Female	-0.0125	-0.0124	-0.0123	-0.00939
	(0.0141)	(0.0141)	(0.0140)	(0.0143)
Publications$_{t-1}$		0.00000305	0.0000238	0.0000262
		(0.0000284)	(0.0000479)	(0.0000458)
Citations$_{t-1}$			-0.0000762	-0.000165
			(0.000136)	(0.000148)
Grants$_{t-1}$				7.63e-11
				(4.84e-11)
Sector FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Observations	443	443	443	443
R-sq.	0.0624	0.0625	0.0630	0.0723

Notes: This table shows the determinants of founders' reliance on their previous academic work when creating their startup. There are 407 unique founders but some of them are associated with more than one startup. The dependent variable is the percentage of scientific citations in a startup's (first granted) patents that come from papers written by founders themselves. *Publications*$_{t-1}$ is the aggregate number of publications published before startup creation. *Citations*$_{t-1}$ is the aggregate number of (year-adjusted) citations received before startup creation. *Grants*$_{t-1}$ is the total amount of grants received before startup creation. We add sector and startup creation year fixed effects in each model. There are more observations that the number of unique professors because some of them are associated with several startups. Standard errors (in parentheses) are clustered at the founder level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: OLS Performance Outcomes, Firm Level

	(1) $\mathbf{1}\{\text{Funds} > p(75)\}$	(2) Acquired	(3) IPO
Percentage self-cites	-0.413**	-0.275***	0.0943
	(0.175)	(0.101)	(0.183)
Number patents (log)	-0.0543	-0.0436	0.0415
	(0.0712)	(0.0447)	(0.0759)
Number scient. patents (log)	0.105	0.0426	0.0126
	(0.0694)	(0.0442)	(0.0692)
Team size (log)	0.0250	0.0607	-0.0176
	(0.0968)	(0.0542)	(0.0567)
At least one female founder	0.0230	-0.0896***	0.00250
	(0.0774)	(0.0326)	(0.0443)
At least one top-tier university	0.117*	0.0137	0.0281
	(0.0598)	(0.0358)	(0.0412)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	308	308	308
R-sq.	0.202	0.168	0.154

Notes: Each observation corresponds to a startup. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than $10million (representing the 75^{th} percentile of the distribution) within the first 5 years of inception. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: 2SLS Performance Outcomes, Firm Level

	(1) $\mathbf{1}\{\text{Funds} > p(75)\}$	(2) Acquired	(3) IPO
Percentage self-cites	-1.983*	-4.254**	-1.531
	(1.092)	(1.601)	(1.208)
Founder controls	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	301	301	301
F-Stat	28.2	28.2	28.2

Notes: This table shows the result of our 2SLS estimation using the number of unique co-authors affiliated with an academic institution before entering entrepreneurship as an instrument. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than $10million (representing the 75^{th} percentile of the distribution) within the first 5 years of inception. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we control for the number of publications published before startup creation, an indicator equal to 1 if at least one founder has a publication co-authored with an individual affiliated with a firm, an indicator equal to 1 if at least one founder has patented before entering entrepreneurship, an indicator equal to 1 if there is at least one female in the founding team, an indicator equal to 1 if at least one founder graduated from a top-tier university and an indicator equal to 1 if at least one founder is a full professor. We also include state, founding year and sector fixed effects. We cluster standard errors at the state and startup level. We report the F-statistic of the first-stage in the last row.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: OLS Innovation Outcomes, Firm Level

	(1) Nb patents (log)	(2) Nb patents (Poisson)
Percentage self-cites	-0.241	-0.106
	(0.315)	(0.512)
Team size (log)	-0.0120	-0.0178
	(0.167)	(0.243)
At least one female founder	-0.189	-0.301*
	(0.125)	(0.172)
At least one top-tier university	-0.0295	0.0237
	(0.108)	(0.167)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	308	308
R-sq.	0.203	

Notes: Each observation corresponds to a startup. The outcome is the aggregate number of U.S. granted patents of a startup. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Forward Citations Patent Level

	(1)	(2)	(3)	(4)
	Forward citations			
	First Patents		All Patents	
Percentage self-cites	0.286	0.662	-0.368	-0.125
	(0.518)	(0.447)	(0.513)	(0.431)
Application Year Controls	Yes	Yes	Yes	Yes
Sector Control	Yes	Yes	Yes	Yes
Patent Class Control	No	Yes	No	Yes
Observations	409	405	1,623	1,596

Notes: The outcome variable is the number of forward citations received by patents. Columns (1) and (2) include only the first patents associated with a startup, while columns (3) and (4) consider all patents associated with a startup. In each model, we control for the number of publications published before startup creation and we add patent application year and sector fixed effects. Columns (2) and (4) add patent class fixed effects. All models are estimated with a Poisson specification. We cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Characteristics of the science cited by startup's patents, Firm Level

	(1)	(2)	(3)
	Concepts		Cites from patents
	Number (log)	Use in other papers (log)	(Poisson)
Percentage self-cites	-0.596**	-0.765**	-1.436*
	(0.269)	(0.365)	(0.820)
Number patents (log)	-0.0763	-0.468**	-0.401
	(0.0962)	(0.228)	(0.303)
Number scient. patents (log)	0.104	0.428*	0.198
	(0.103)	(0.237)	(0.306)
Team size (log)	0.0367	0.0460	-0.417
	(0.0611)	(0.132)	(0.282)
At least one female founder	0.0151	0.0958	-0.137
	(0.0493)	(0.112)	(0.249)
At least one top-tier university	-0.0406	0.0151	0.146
	(0.0390)	(0.0923)	(0.183)
State Controls	Yes	Yes	Yes
Year Controls	Yes	Yes	Yes
Publication Year Controls	Yes	Yes	Yes
Observations	282	282	287
R-sq.	0.242	0.352	

Notes: Each observation corresponds to a startup. In column (1), the outcome is the (log) average number of concepts used in the papers cited by the first patents of a startup. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. In column (3), the outcome is the number of citations received from patents (excluding the ones from the focal startup) that the papers have received. In each model, we control for the number of publications published before startup creation, we add sector, state, startup creation year and paper publication year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Summary statistics for the treatment and control groups of the difference-in-difference analysis

	Control group % self-cites= 0	Treatment group % self-cites> 0	p-value of difference
Panel A: Time invariant characteristics			
Startup's number of patents	6.3	6.4	0.92
Startup's number of patents in RoS	6.0	6.0	0.91
Female	0.09	0.08	0.69
Top-tiers university	0.41	0.44	0.61
Biotechnology sector	0.82	0.86	0.52
Panel B: At time of startup creation			
Academic experience	21.3	22.2	0.46
Publications (cumulative)	138	133-7	0.94
Top publications (cumulative)	61	54	0.39
Citations ('000s) (cumulative)	16.9	13.0	0.21
% papers last author (average)	38.8	38.8	0.97
% papers first author (average)	24.7	25.2	0.77
Number of unique individuals	250	125	

Cumulative: values of the variable are summed from career start year to startup creation year.

Average: values of the variable from career start year to startup creation year are averaged.

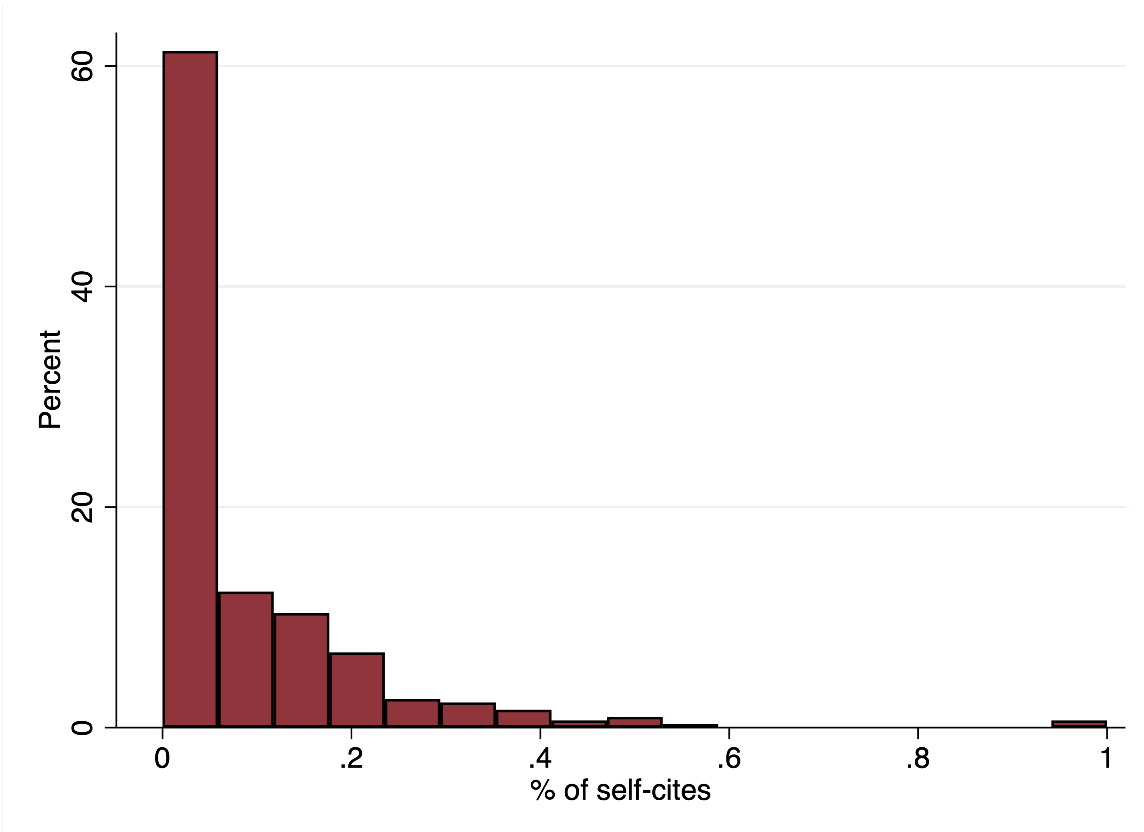
Table 9: Difference-in-difference analysis of academic output

	(1) Publications (log)	(2) Top (log)	(3) Cites (log)	(4) % first (log)	(5) % last (log)
$\text{Treated}_i \times \text{Post}_t$	-0.0279	-0.00368	-0.0222	0.0121	0.0127
	(0.0679)	(0.0514)	(0.0507)	(0.0104)	(0.0151)
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes
Observations	13,345	7,091	7,091	11,986	11,986
R-sq.	0.547	0.630	0.430	0.347	0.318

Notes: In column (1), the outcome is the number of publications published by founders in a specific year. In column (2), the outcome is the number of top publications, defined as publications which received more citations than other articles published the same year in the same area of research. In column (3), the outcome is the average number of cites (adjusted for field and publication year) received by articles published in a specific year. In column (4), the outcome is the share of papers where the founder appears as first author. In column (5), the outcome is the share of papers where the founder appears as the last author. All outcomes are logged. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

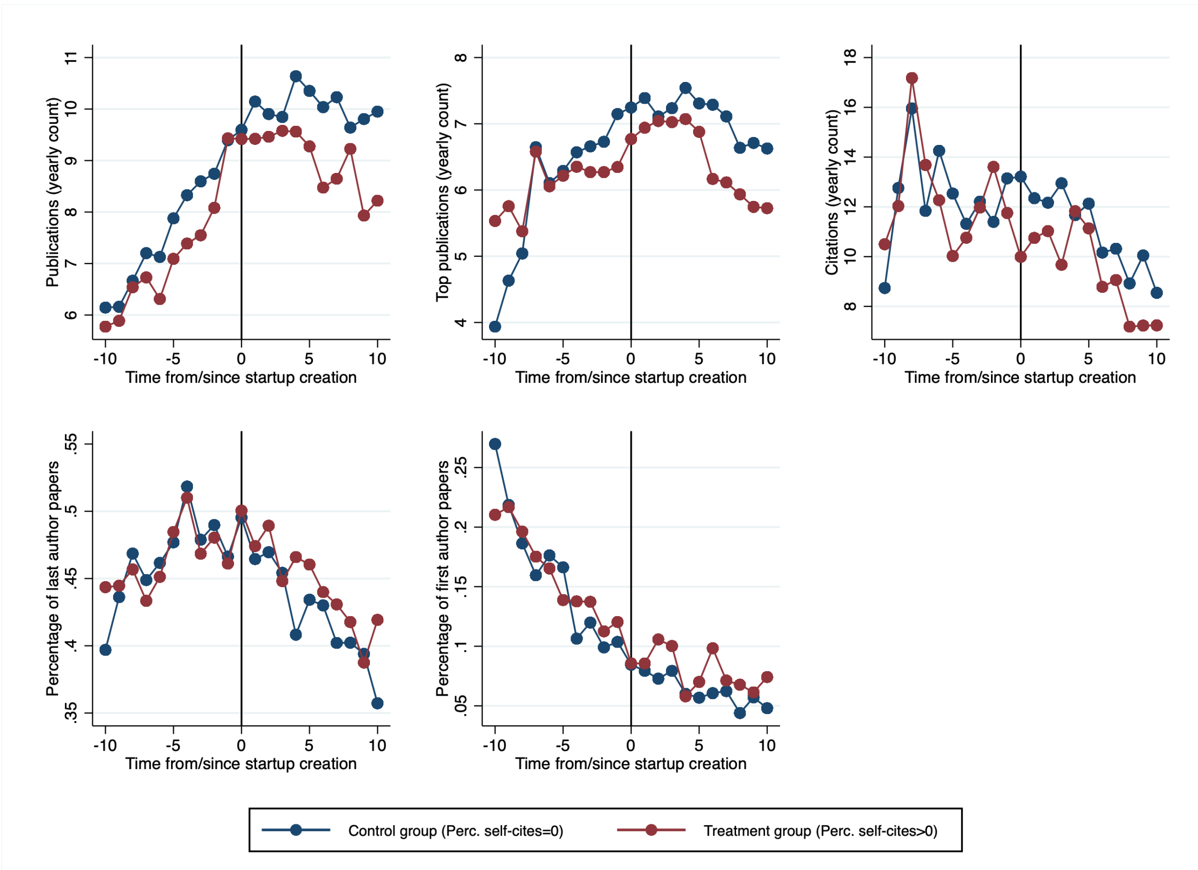
Online Appendix for: Bringing Science to Market: Knowledge Foundations and Performance

Figure A1: Histogram of the treatment variable



Notes: This figure shows the distribution of the *Percentage self-cites* variable. To calculate this variable, we calculate the percentage of citations that the first granted patents of a startup make to its founders' academic papers. We then average this measure at the startup level.

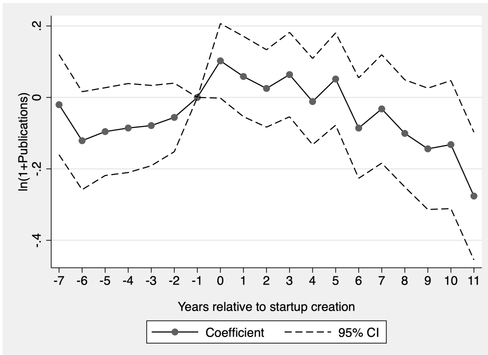
Figure A2: Raw plot of the outcomes for the treatment and control groups



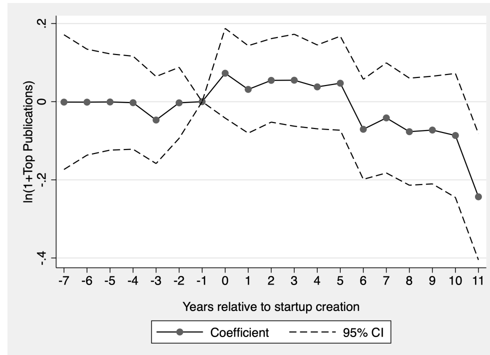
Notes: This figure shows the mean values before startup creation (negative values of the x-axis) and after startup creation (positive values of the x-axis) of several outcome variables for the treatment (red line) and control (blue line) groups. The treatment group is composed of individuals whose startup has a patent that cites at least one of their academic papers (i.e., *Percentage self-cites* > 0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites* $= 0$). Outcomes are (from left to right, top to bottom): the log yearly number of publications, the log yearly number of top publications, the log yearly number of (year- and field-adjusted) citations, the log yearly number of publications where the founder appears as first author and the log yearly number of publications where the founder appears as last author.

Figure A3: Event study graphs

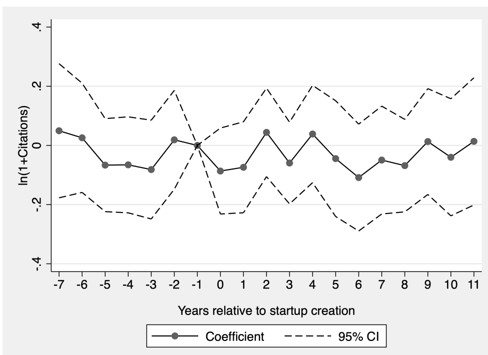
(a) Log(1+Publications)



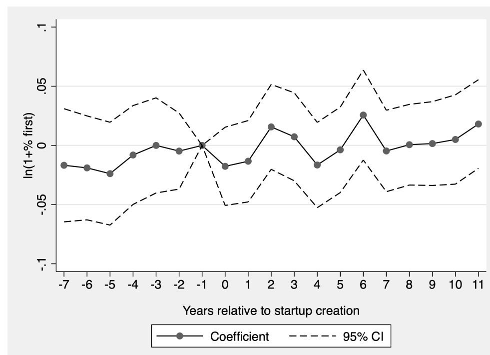
(b) Log(1+Top Publications)



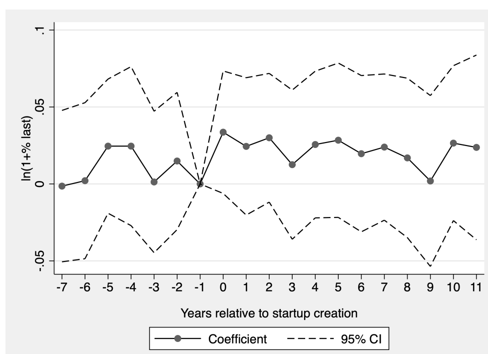
(c) Log(1+Citations)



(d) Log(1+% papers as first author)

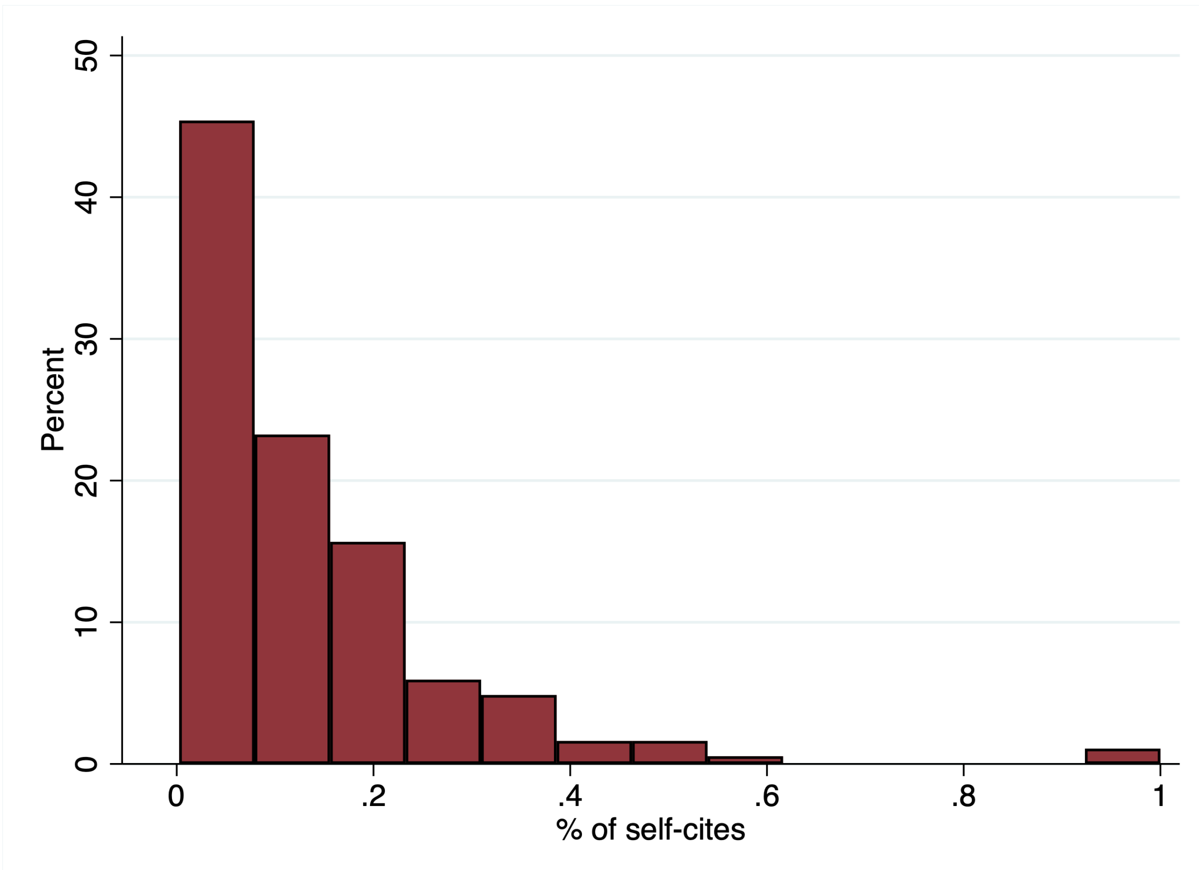


(e) Log(1+% papers as last author)



Notes: This figure shows the difference-in-difference coefficients over time, before startup creation (negative values of the x-axis) and after startup creation (positive values of the x-axis). We normalize all coefficients with respect to the year preceding startup creation. The treatment group is composed of individuals whose startup has a patent that cites at least one of their academic papers (i.e., *Percentage self-cites* > 0), while the control group is composed of individuals whose startup has no patent citing their academic work (*Percentage self-cites* $= 0$). Outcomes are the log yearly number of publications (panel (a)), the log yearly number of top publications (panel (b)), the log yearly number of (field and publication-year adjusted) citations (panel (c)), the log yearly number of publications where the founder appears as first author (panel (d)) and the log yearly number of publications where the founder appears as last author (panel (e)). Dotted lines represent the 95% confidence interval.

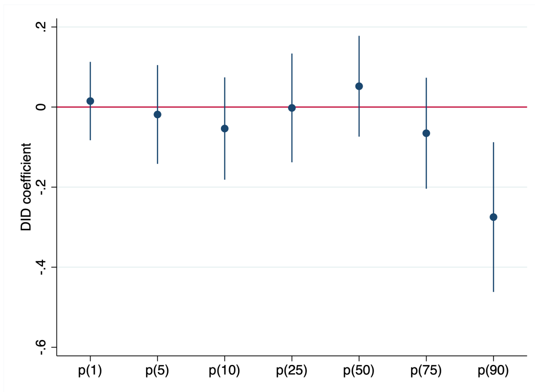
Figure A4: Histogram of the treatment variable for startups that rely on Science



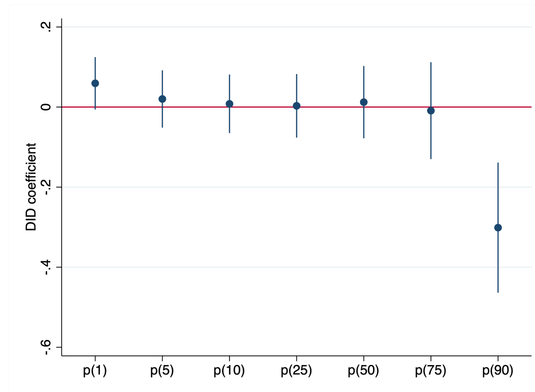
Notes: This figure shows the distribution of the *Percentage self-cites* variable when restricted to startups that cite at least one of their founders' academic work (*Percentage of self-cites*> 0).

Figure A5: Heterogeneity Analysis of Academic Output, startups with *Percentage self-cites*> 0

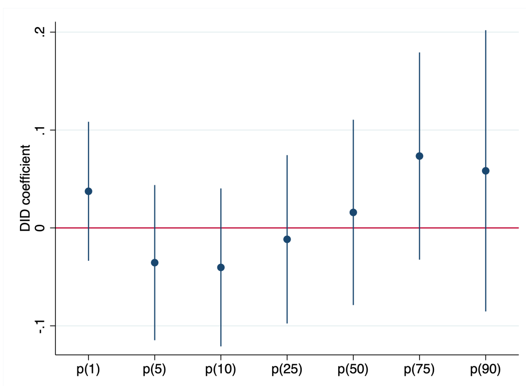
(a) Log(1+Publications)



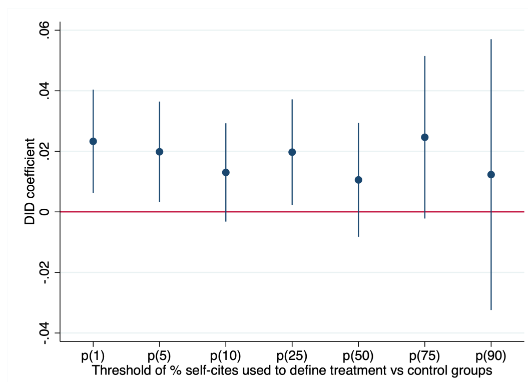
(b) Log(1+Top Publications)



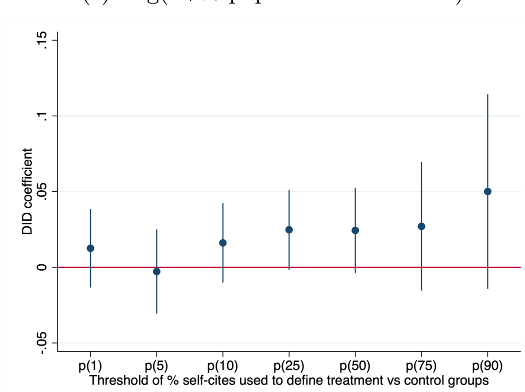
(c) Log(1+Citations)



(d) Log(1+% papers as first author)

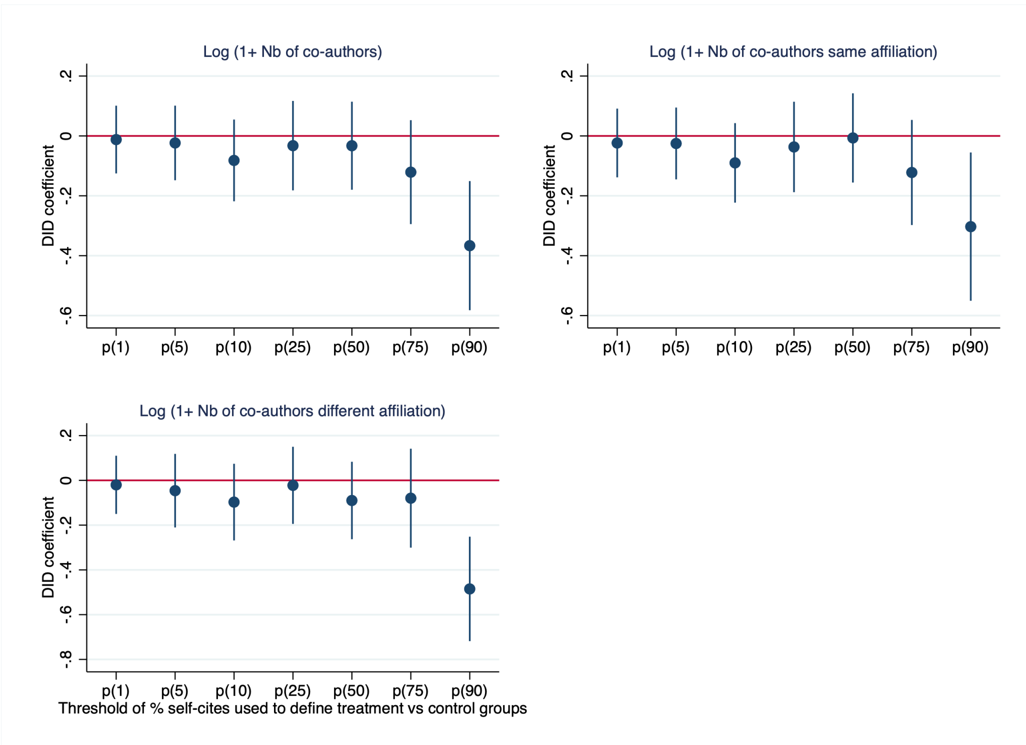


(e) Log(1+% papers as last author)



Notes: This figure shows the difference-in-difference coefficients using different thresholds for defining the treatment and control groups. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. For instance, p(1) means that all individuals whose *Percentage self-cites* variable is below the 1^{st} percentile are part of the control group, while individuals whose *Percentage self-cites* variable is above the 1^{st} percentile are part of the treatment group. In Panel (a), the outcome is the log yearly number of publications. In Panel (b), the outcome is the log yearly number of top publications. In Panel (c), the outcome is the log average number of citations. In Panel (d), the outcome is the log yearly number of publications where the founder appears as first author. In Panel (e), the outcome is the log yearly number of publications where the founder appears as last author. Each bar denotes the 90% confidence interval.

Figure A6: Heterogeneity Analysis of the Number of Co-Authors, startups with *Percentage self-cites*> 0

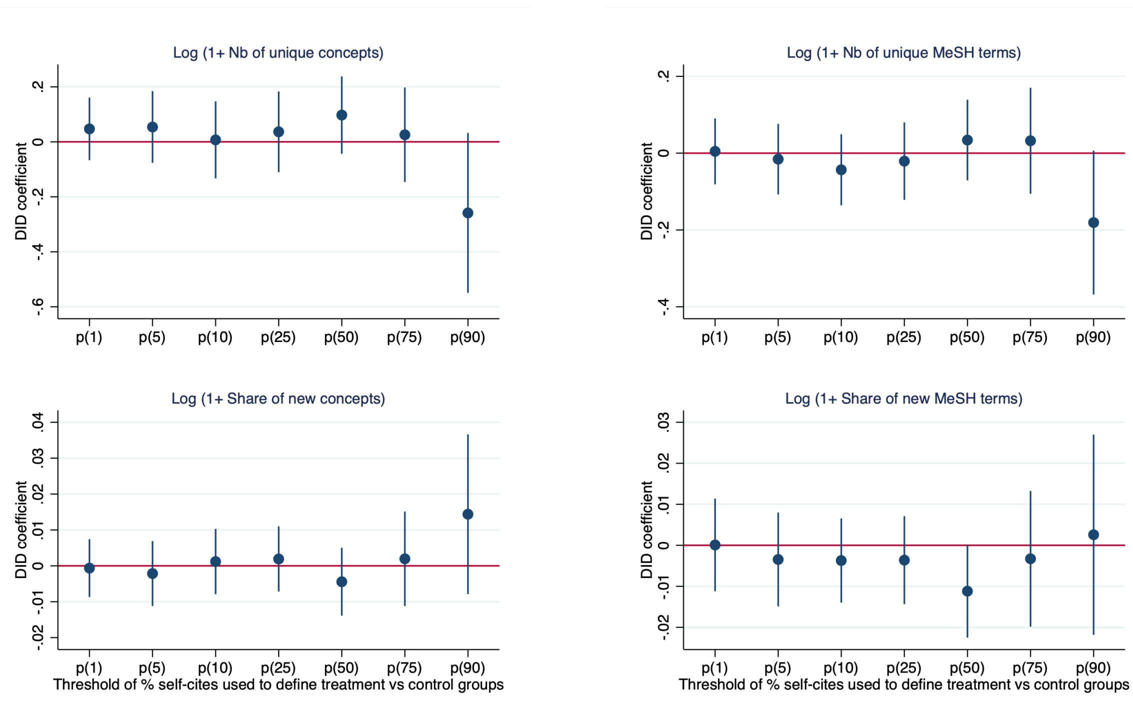


Notes: This figure shows the difference-in-difference coefficients. The outcomes are the log number of unique co-authors founders published with in a given year (top left figure), the log number of unique co-authors from the same institution founders published with in a given year (top right figure) and the log number of unique co-authors from a different institution founders published with in a given year (bottom figure). The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. In the first row, t Each bar denotes the 90% confidence interval.

Figure A7: Heterogeneity Analysis of Research Focus, using Dimensions AI concepts, startups with *Percentage self-cites*> 0

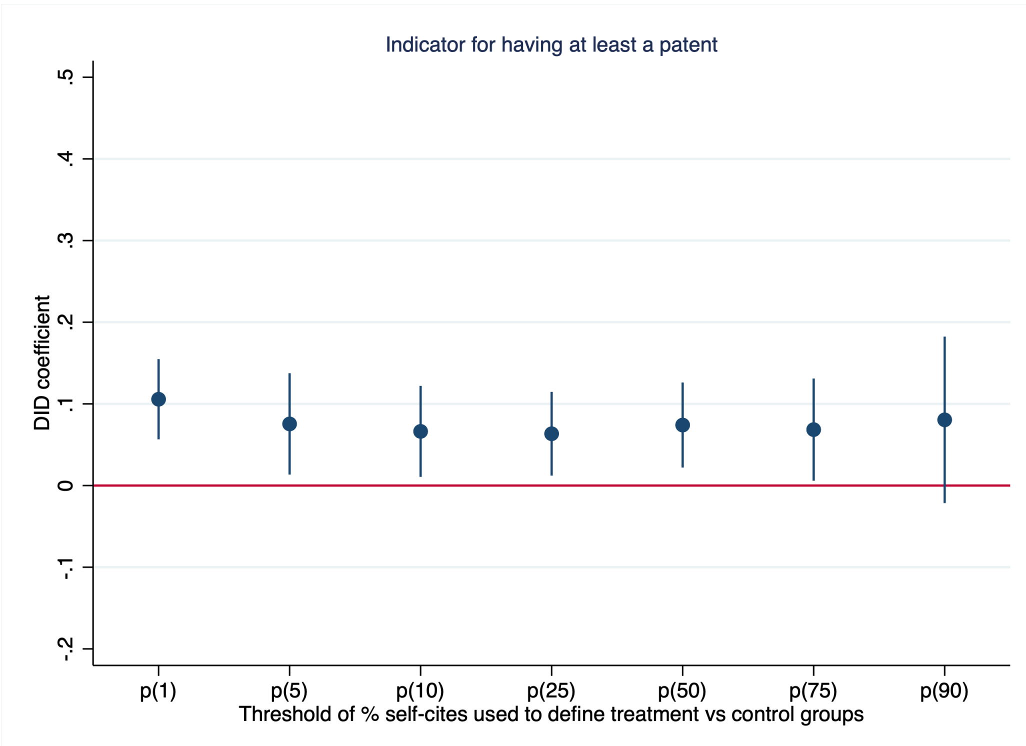
(a) Using the same control group and varying the treatment group

(b) Varying the threshold to define treatment and control groups



Notes: This figure shows the difference-in-difference coefficients. In panel (a), the outcome is the log number of unique concepts founders used in a given year in their papers (top row) and the log number of the share of new concepts founders used in a given year in their papers (bottom row). In panel (b), the outcome is the log number of unique MeSH terms founders used in a given year in their papers (top row) and the log number of the share of new MeSH founders used in a given year in their papers (bottom row). The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. Each bar denotes the 90% confidence interval.

Figure A8: Heterogeneity Analysis, Patenting Activity, startups with *Percentage self-cites*> 0



Notes: This figure shows the difference-in-difference coefficients. The outcome is an indicator variable equal to 1 if the founder filed a patent in a given year and 0 otherwise. The x-axis shows the threshold for defining the control (below the threshold) and the treatment (above the threshold) groups. Each bar denotes the 90% confidence interval.

Table A1: OLS Stage of Development Outcomes, Firm Level

	Time from paper to patent		Time from paper to startup	
	(1)	(2)	(3)	(4)
Percentage self-cites	-13.37*	-10.93	-8.950	-4.648
	(7.009)	(7.628)	(6.664)	(7.048)
Team size (log)		2.682		3.155
		(2.805)		(2.973)
At least one female founder		-2.517		-0.843
		(2.382)		(2.449)
At least one top-tier university		-1.648		-1.812
		(2.046)		(2.070)
State FE	No	Yes	No	Yes
Founding Year FE	No	Yes	No	Yes
Sector FE	No	Yes	No	Yes
Observations	288	288	288	288
R-sq.	0.0126	0.159	0.00564	0.146

Notes: This table shows the correlation between *Percentage of self-cites* and proxies of the time needed to bring scientific knowledge from Academia to the private sector. Columns (1) and (2) use the time between a startup's first patent application year and its first scientific cite. Columns (3) and (4) use the time between a startup's creation year and the first scientific cite. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Difference-in-difference analysis of number of co-authors by affiliation

	(1) All co-authors (log)	(2) Same affiliation (log)	(3) Different affiliation (log)
$\text{Treated}_i \times \text{Post}_t$	-0.0556	-0.0511	-0.0958
	(0.0770)	(0.0782)	(0.0897)
Year FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes
Observations	11,853	11,853	11,853
R-sq.	0.638	0.630	0.484

Notes: In column (1), the outcome is the (log) number of unique co-authors each founder has in a given year. In column (2), the outcome is the (log) number of unique co-authors from the same institution that each founder has in a given year. In column (3), the outcome is the (log) number of unique co-authors from a different institution that each founder has in a given year. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Difference-in-difference analysis of academic output - Robustness check

	(1) Publications (log)	(2) Top (log)	(3) Cites (log)	(4) % first (log)	(5) % last (log)
$\text{Treated}_i \times \text{Post}_t$	-0.00500	0.0246	-0.0320	0.00275	0.00500
	(0.0666)	(0.0467)	(0.0462)	(0.00964)	(0.0136)
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes
Observations	64,710	32,064	32,064	57,611	57,611
R-sq.	0.557	0.638	0.413	0.361	0.325

Notes: In this specification, we follow Goodman-Bacon (2021) and exclude the post-period of earlier treated groups when using them as control groups. In column (1), the outcome is the number of publications published by founders in a specific year. In column (2), the outcome is the number of top publications, defined as publications which received more citations than other articles in the same area of research. In column (3), the outcome is the cumulative number of cites received by articles published in a specific year. In column (4), the outcome is the share of papers where the founder appears as first author. In column (5), the outcome is the share of papers where the founder appears as the last author. In each model, we add publication year, individual and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: OLS Performance Outcomes, *Percentage self-cites*>0, Firm Level

	(1) $\mathbf{1}\{\text{Funds} > p(75)\}$	(2) Acquired	(3) IPO
Percentage self-cites	-0.408*	-0.213*	0.142
	(0.217)	(0.111)	(0.208)
Number patents (log)	-0.170**	-0.0533	0.0773
	(0.0678)	(0.0368)	(0.139)
Number scient. patents (log)	0.212***	0.0361	-0.0115
	(0.0784)	(0.0444)	(0.142)
Team size (log)	0.0229	0.0467	-0.0431
	(0.112)	(0.0612)	(0.0752)
At least one female founder	0.0609	-0.0795**	-0.0217
	(0.0960)	(0.0357)	(0.0530)
At least one top-tier university	0.197**	0.0325	0.0497
	(0.0839)	(0.0412)	(0.0625)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	185	185	185
R-sq.	0.259	0.245	0.191

Notes: Each observation corresponds to a startup. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than $10million (representing the 75^{th} percentile of the distribution) within the first 5 years of inception. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: OLS Innovation Outcomes, *Percentage self-cites* > 0, Firm Level

	(1) Nb patents (log)	(2) Nb patents (Poisson)
Percentage self-cites	-0.522	-0.765
	(0.399)	(0.562)
Team size (log)	0.00492	-0.0350
	(0.214)	(0.274)
At least one female founder	-0.157	-0.264
	(0.177)	(0.221)
At least one top-tier university	0.0432	0.197
	(0.153)	(0.187)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	185	185
R-sq.	0.225	

Notes: Each observation corresponds to a startup. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Characteristics of the science cited by startup's patents, *Percentage self-cites* > 0, Firm Level

	(1)	(2)	(3)
	Concepts		Cites from patents
	Number (log)	Use in other papers (log)	(Poisson)
Percentage self-cites	-0.655*	-1.383***	-2.334**
	(0.366)	(0.377)	(0.983)
Number patents (log)	0.000622	-0.291**	-0.412
	(0.0614)	(0.138)	(0.338)
Number scient. patents (log)	-0.00842	0.271*	0.198
	(0.0665)	(0.156)	(0.373)
Team size (log)	0.0431	-0.0422	-0.188
	(0.0678)	(0.148)	(0.268)
At least one female founder	0.0335	0.134	0.245
	(0.0494)	(0.128)	(0.258)
At least one top-tier university	0.000651	-0.00141	0.0799
	(0.0418)	(0.104)	(0.206)
State Controls	Yes	Yes	Yes
Year Controls	Yes	Yes	Yes
Publication Year Controls	Yes	Yes	Yes
Observations	181	181	182
R-sq.	0.508	0.455	

Notes: Each observation corresponds to a startup. In column (1), the outcome is the (log) average number of concepts used in the papers cited by the first patents of a startup. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. In column (3), the outcome is the number of citations received from patents (excluding the ones from the focal startup) that the papers have received. In each model, we control for the number of publications published before startup creation, we add sector, state, startup creation year and paper publication year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: OLS Performance Outcomes, Firm Level, All Patents

	(1) $\mathbf{1}\{\text{Funds} > p(75)\}$	(2) Acquired	(3) IPO
Percentage self-cites	-0.597***	-0.256**	-0.0879
	(0.211)	(0.124)	(0.153)
Number patents (log)	-0.0596	-0.0383	0.0288
	(0.0716)	(0.0443)	(0.0768)
Number scient. patents (log)	0.110	0.0378	0.0246
	(0.0699)	(0.0437)	(0.0703)
Team size (log)	0.0283	0.0664	-0.0238
	(0.0954)	(0.0543)	(0.0568)
At least one female founder	0.0226	-0.0879***	-0.000398
	(0.0770)	(0.0322)	(0.0439)
At least one top-tier university	0.112*	0.0124	0.0261
	(0.0595)	(0.0360)	(0.0409)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	308	308	308
R-sq.	0.207	0.162	0.153

Notes: Each observation corresponds to a startup. We use all patents pertaining to a startup to calculate the treatment variable. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than \$10million (representing the 75^{th} percentile of the distribution) within the first 5 years of inception. In column (2), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (3), the outcome *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: OLS Innovation Outcomes, Firm Level, All Patents

	(1) Nb patents (log)	(2) Nb patents (Poisson)
Percentage self-cites	-0.258	-0.394
	(0.413)	(0.585)
Team size (log)	-0.00780	-0.0278
	(0.167)	(0.248)
At least one female founder	-0.188	-0.298*
	(0.125)	(0.173)
At least one top-tier university	-0.0311	0.0163
	(0.108)	(0.167)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	308	308
R-sq.	0.203	

Notes: Each observation corresponds to a startup. We use all patents pertaining to a startup to calculate the treatment variable. In column (1), the outcome is the aggregate number of U.S. granted patents of a startup (expressed in natural logarithm). In column (2), we fit a Poisson model to the aggregate number of U.S. granted patents of a startup. In each model, we control for the number of publications published before startup creation, we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Forward Citations Patent Level, All Patents

	(1)	(2)	(3)	(4)
	Forward citations (Poisson)			
	First Patents		All Patents	
Percentage self-cites	-1.180*	-0.554	-0.414	-0.0145
	(0.650)	(0.357)	(0.546)	(0.488)
Application Year Controls	Yes	Yes	Yes	Yes
Sector Control	Yes	Yes	Yes	Yes
Patent Class Control	No	Yes	No	Yes
Observations	421	416	1,638	1,610

Notes: The outcome variable is the number of forward citations received by patents. We use all patents pertaining to a startup to calculate the treatment variable. Columns (1) and (2) include only the first patents associated with a startup, while columns (3) and (4) consider all patents associated with a startup. In each model, we control for the number of publications published before startup creation and we add patent application year and sector fixed effects. Columns (2) and (4) add patent class fixed effects. All models are estimated with a Poisson specification. We cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Characteristics of the science cited by startup's patents, Firm Level, All Patents

	(1)	(2)	(3)
	Concepts		Cites from patents
	Number (log)	Use in other papers (log)	(Poisson)
Percentage self-cites	-0.193	-0.861**	-2.768***
	(0.169)	(0.375)	(0.897)
Number patents (log)	-0.0204	-0.322*	-0.475
	(0.0837)	(0.178)	(0.313)
Number scient. patents (log)	0.0513	0.337*	0.273
	(0.0917)	(0.191)	(0.315)
Team size (log)	0.0406	-0.0122	-0.426
	(0.0445)	(0.114)	(0.277)
At least one female founder	0.0379	0.142	-0.110
	(0.0398)	(0.110)	(0.247)
At least one top-tier university	-0.0186	0.0592	0.137
	(0.0293)	(0.0814)	(0.180)
State Controls	Yes	Yes	Yes
Year Controls	Yes	Yes	Yes
Publication Year Controls	Yes	Yes	Yes
Observations	284	284	287
R-sq.	0.197	0.411	

Notes: Each observation corresponds to a startup. We use all patents pertaining to a startup to calculate the treatment variable. In column (1), the outcome is the (log) average number of concepts used in the papers cited by the first patents of a startup. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. In column (3), the outcome is the number of citations received from patents (excluding the ones from the focal startup) that the papers have received. In each model, we control for the number of publications published before startup creation, we add sector, state, startup creation year and paper publication year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: 2SLS Performance Outcomes for Success and IPO, Firm Level

	(1)	(2)	(3)
Dep. Var.	Percentage self-cites	Success	IPO
Model	First Stage	IV	IV
Log(1+Network size)	-0.0578***		
	(0.0173)		
Percentage self-cites		-5.502***	-2.612*
		(1.970)	(1.492)
Founder controls	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
State × Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	326	326	326
F-Stat		11.2	11.2

Notes: This table shows the result of our 2SLS estimation using the number of unique co-authors before entering entrepreneurship as an instrument. Each model includes as controls the log number of patents, the log number of patents relying on scientific literature, the log of team size calculated with the number of founders at inception, an indicator equal to 1 if there is at least one female in the founding team and an indicator equal to 1 if at least one founder graduated from a top-tier university. We also include state, founding year, state × founding year and sector fixed effects. We cluster standard errors at the state and startup level. Results are robust to clustering only at the startup level. Column (1) shows the first-stage regression. Column (2) shows the 2SLS results for *Success*, an indicator variable equal to 1 if the startup is acquired or went public via an IPO. Column (3) shows the 2SLS results for *IPO* is an indicator variable equal to 1 if the startup went public via an IPO. We report the F-statistic of the first-stage in the last row of columns (2) and (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Academic and patenting output using Dimensions AI

In order to identify publications and patents associated with each professor in our sample, we use the Dimensions AI database, which is very similar to Scopus and Web of Science (Singh et al., 2021). Interestingly for us, Dimensions provides a disambiguation of researchers where they take into account a variety of variables such as existing person IDs, name variants, affiliation data, research topics, journals, co-authors, and active years in order to aggregate publications, patents and grants into author profile[23]. Our algorithm relies on identifying the Dimensions researcher IDs associated with each of the professors in our sample, in order to then retrieve information about publications and patents associated with these IDs. Note that several IDs can be associated with one individual. Our algorithm matches each professor with one (or more) researcher ID(s) using the following procedure:

- We first clean institution names in our sample and match each institution to its equivalent in Dimensions. Each institution is therefore associated with a "grid.id" identifier
- We clean first names in the obvious cases where the nickname was used as first name (e.g., some individuals in our sample have first name "Bill" that we transform into "William")
- At this point, each row in our dataset corresponds to an individual for whom we know last name, first name, and grid.id. The goal of the procedure is now to find the researcher ID(s) associated with each row
- For this, for each individual (or equivalently row), we perform an exact match based on last name and grid.id. This gives us all the potential researcher ID(s) that could be a match because they worked at the same institution as our focal professor at one point during their career and have an identical last name. We do not perform an exact match on the first name as this point as Dimensions may combine first and middle names into first name, which would make us miss potential matches when our main dataset does not include middle name
- For each individual, we search among his potential matches and keep only those whose associated first name in Dimensions includes the first name of our focal professor. For example, if our individual has first name "Carl" and we have 3 potential matches for this individual whose first names are respectively "Carl K", "Tom" and "Carter", only the first match will be kept
- Among the individuals for which the previous procedure led to no match, we perform a fuzzy match on first name. This is useful for individuals for which the first-name in our dataset has an hyphen (for instance "kwok-kin" should be matched to the Dimension researcher "kwok kin" but this was missed in the previous step. Similarly, "Robert" is a match for the first name "Rob" but would have been missed otherwise). We test several thresholds for the fuzzy match and end up selecting matches whose score is above 80. We manually check every match to ensure accuracy
- We exclude 2 individuals for which there was a high number of researcher IDs matches (11 and 12 matches respectively)
- For researchers where there is no middle name in our dataset, we manually search online to find it and keep researcher IDs matches whose middle name is the same as the one identified

This procedure leads us with 561 professors for whom we have identified one or more researcher IDs. Given that a professor can be associated with several researcher IDs, we then collapse publications, citations and patents at the professor level.

[23]https://dimensions.freshdesk.com/support/solutions/articles/23000018779-how-are-researchers-unified-disambiguated-in-dimensions-