Bringing Science to Market: Knowledge Foundations and Performance*

Justine Boudou
Harvard Business School

Maria Roche
Harvard Business School

May 7, 2024

Abstract

Possessing unique knowledge is widely considered a critical source of competitive advantage. In this paper, we examine the relationship between the extent to which founders exploit their own technologically unique knowledge and subsequent new venture performance. Using a panel dataset of 510 academic startups in bio-medicine 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 less likely to receive higher levels of funding, while we do not observe differences in their likelihood to go through an IPO. We provide evidence suggesting that the underlying mechanism driving these results may be rooted in the *nature* of the scientific knowledge these startups rely on: more narrow and specialized.

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, DRUID and the Productivity Seminar. We are indebted to Ashish Arora, George Chondrakis, Shane Greenstein, Jacqueline Lane, Brian Silverman, and Keyvan Vakili for feedback on early drafts. We are grateful for the valuable insights provided by several technology transfer offices and entrepreneurial faculty. 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 notion 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 (Geroski et al., 2010; Park et al., 2023).

In this paper, we examine the relationship between the extent to which startups exploit their founders' technologically unique knowledge and new venture performance outcomes. We thereby pay particular attention to the specific nature of knowledge these founders incorporate in their ventures. The context of our study is 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 (Bercovitz and Feldman, 2006). 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). This situation has raised concerns that university-based knowledge stays stuck in an "ivory tower",

[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).

not sufficiently diffusing out into the economy (Matthews, 2023). Although some studies have highlighted that heterogeneity among 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 (e.g., Barney 1991), 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 bio-medicine established 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 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 characterize founders' academic work, we consider their published papers before startup creation. To capture the initial knowledge-base of a startup, we look at its first granted patents, at or close to the time of creation. In the next steps, to operationalize the distance variable, we 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

[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.

and performance outcomes we take a step-wise 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.

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 10 percentage-point increase in founders' use of their previous academic work when creating their startup is associated with a 2.2% lower probability of acquisition and a 3.4% 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 then explore potential mechanisms underlying our results. We provide suggestive evidence that these results do not appear to be driven by lower invention quality nor a more nascent technology. Similarly, when examining founders' involvement in their ventures, we do not find that differences in the degree to which founders take an active management role are the main driver of our results. Instead, upon closer examination of startups' knowledge-base, we find that variation in the *nature* of the scientific knowledge utilized appears to be the most feasible mechanism. 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, a feature highly valued in academia, 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, and less flexible to change (pivot) as market demand becomes more clear (Roche et al., 2020). This reduces an inventions' appeal for licensing (Lee, 2023), but also as an acquisition target (Arora, Belenzon and Suh, 2022; Polidoro 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 other use cases, scalability and market impact, in line with the lower probability of raising a significant amount of funds (Aggarwal et al., 2015). 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 it 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 (Bayar and Chemmanur, 2011), rather than the specific intricacies of the technology.

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 breadth of the knowledge that they use. In particular, our findings stress the importance of considering the *nature* of scientific knowledge (i.e., how narrow and specialized to the founder) as a strategic choice when founding and developing startups. 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). Our results underscore the multifaceted nature of strategic decisions made by academic founders, carrying potentially critical strategic and practical implications for their ventures.

2 Conceptual framework

2.1 Academic entrepreneurship and knowledge endowments

Scientific knowledge has emerged as a powerful catalyst for innovation. For example, patents that rely on scientific content have been found to generate more follow-on citations (Sorenson and Fleming, 2004), to more likely be renewed (Ahmadpoor and Jones, 2017), to generate more value for firms (Krieger et al., 2022) and to more likely 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 U.S. since the Bayh-Dole Act of 1980 (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.

As such, academic startups 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 (Aghion et al., 2008). However, despite their potential, academic startups do not perform particularly well on the market on average (Park et al., 2023; Roche et al., 2020), calling for a deeper examination of the mechanisms that underpin their performance.

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 formed 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; Roach, 2017; Roach and Sauermann, 2015; 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, the influence of *initial technical knowledge endowments* on subsequent firm performance remains understudied. In particular, while it is widely acknowledged that science is a potential driver of innovation, the distinct characteristics and implications associated with the specific *nature* of science employed in entrepreneurial ventures remain largely unexplored.

Variation in knowledge endowments may be particularly salient in the context of academic entrepreneurship. Academic entrepreneurs possess specialized, often tacit knowledge (Bercovitz and Feldman, 2006) in a narrow field of science that is unique to their own research and expertise (Jones, 2009). This specialized knowledge, derived from years of academic training and research, grant 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). By having more intimate knowledge about the technology surrounding their startup, founders who rely more heavily on their own academic work may be better equipped to identify when and why a technology might succeed or fail (Kacperczyk and Younkin, 2017), giving their startup a competitive advantage in terms of technical expertise and innovation potential. As a consequence, academic startups that more extensively leverage their founders' unique technical knowledge may be better positioned to develop groundbreaking innovations and capture market opportunities that are inaccessible to others.

Furthermore, the dual involvement of academic founders in both academia and entrepreneurship may present opportunities for synergy and cross-fertilization of ideas (Bikard et al., 2019). Founders who rely more on their own academic work might be able to more easily access the necessary equipment and instruments needed to develop the technology through their own laboratories, facilitating the transfer of their cutting-edge research into real-world applications (Colombo and Piva, 2012). In addition, founders who rely more on their own academic work may benefit from dynamic interactions between their academic and commercial activities, ultimately enhancing their startup's competitive advantage and potential for success. From this, we may expect *academic startups that are founded based primarily on the scientific work of their founder(s) to be particularly successful*, provided their advantage rooted in knowledge about a specific technology.

2.3 A knowledge-based disadvantage

While specialized knowledge provides clear advantages, its narrow and highly specific nature, often unique to individual founders, can present challenges for startups. Specifically, venture capitalists (VCs) and potential acquirers may face difficulties in comprehending and evaluating the commercial potential of startups rooted in highly specialized scientific domains. For example, the intricate nature of specialized scientific knowledge may hinder communication and collaboration between founders and external stakeholders, potentially limiting access to crucial funding and partnership opportunities (Polidoro and Yang, 2021; Stuart et al., 1999). Moreover, the integration of highly specialized technologies into existing product lines or business models may present challenges for potential acquirers, deterring acquisition or investment in startups relying on such specialized knowledge (Makri et al., 2010). As a consequence, this might lead to lower performance outcomes for startups that rely more on the scientific knowledge of their founders.

In addition to the consequences related to the nature of science they use in their startup, founders who rely more on their own academic work might be differently impacted by the dual involvement in both academia and entrepreneurship (Roche, 2023). Work examining academics' commitment to both academia and commercialization activities has primarily focused on implications for academic output (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).

Evidence regarding the impact of founders' dual involvement on commercial *success* is more limited. Work by Roche et al. (2020) highlights poorer entrepreneurial performance of academic startups though the precise mechanism remains suggestive. Importantly, founders who lean heavily on their own academic work to establish their startup may prefer to retain roles that afford them greater control over the day-to-day operations of their ventures, as opposed to handing over control to someone solely tasked with leading the startup. However, studies suggest that founder control is often linked with poorer venture outcomes (Wasserman, 2017), and this insistence on control may also create conflicts with their academic commitments. As a consequence, academic founders with executive roles may find themselves with limited time to allocate to their ventures, hence jeopardizing their success (Ding and Choi, 2011; Staw, 1981).[4] Additionally, founders who rely more on their scientific knowledge might have a particular attachment to their startup or a particular belief in the viability of their ideas that make them more likely to experience an escalation of commitment (Schmidt and Calantone, 2002), wherein they persist in pursuing failing strategies or ventures. From this, we may expect *lower performance outcomes for startups that rely more on the scientific knowledge of their founder(s).*

Overall, provided the state of extant literature, what the relationship between startups' performance outcomes and their reliance on the scientific knowledge of their founders could be remains

[4]We think as executive roles as positions that require a significant time commitment from founders, such as Chief Executive Officer (CEO), Chief Financial Officer (CFO), Chairman or other similar C"X"O postitions.

ambiguous. In other words, it is *unclear, ex-ante, if the extent to which an academic entrepreneur exploits their own unique technological knowledge impacts their startup success and what the directionality of the relationship may be.* This 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 U.S. 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 bio-medicine (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) (Conti and Roche, 2021). 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 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 further restricts the analysis to startups with at least one patent and for which we can observe pre-entrepreneurship academic output produced by at least one academic founder. This leads to a final sample of 308 academic startups. 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 that 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 taking the pool of papers academics 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 which it builds on, including scientific papers. This allows us to differentiate between citations that patents make to 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 these patents make to their founders' academic papers.

In practice, we take the first granted patent(s) a startup applied for[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.[6] For each academic article cited by a patent, we create a self-citation dummy equal to 1 when at least one author of the academic article is matched to an inventor with a confidence score above 50.[7] This identifies instances where we can

[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.

[6]As detailed in previous work, patents may cite scientific research on the front-page or in the body of the text. Front-page citations are usually aimed at citing prior art while in-text citations are closer to the role played by academic citations, incorporating knowledge by reference (Bryan et al., 2020). Most research so far focused on front-page citations because they were easier to extract. Because the RoS dataset contains both types of citations, our analysis considers both front-page and in-text citations.

[7]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.

reasonably be confident that at least one inventor of the patent is citing their own academic work. 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.[8]

3.3 Summary statistics

Our main independent variable of interest is skewed, with an average value of 8.1% (Figure 1 presents the histogram of *Percentage Self-Cites*). 40% of startups have patents that do not rely on their founders' academic work: our main variable of interest takes a value of 0, implying a relative high distance between the knowledge-base of the startup and that of its founders' academic work.[9]

<Insert Figure 1 here>

Table 1 displays summary statistics for our sample of 308 firms. Startups have on average 7 patents, and on average 6 of these patents 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

[8]One concern could be that our measure captures narcissism rather than knowledge-base proximity. 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 her 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 main independent 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 our main variable of interest 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.

[9]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.

70% have at least one founder who graduated from a top-tier university.[10] 8% of the startups in our sample are acquired and 7% of them go public through an IPO. The amount of funds raised is skewed, with an average of $U.S. 6.7 millions.

<Insert Table 1 here>

Startups in our dataset are primarily located in California (24%), Massachusetts (21%), Pennsylvania (5%), North Carolina (4%), Maryland (4%) and Michigan (4%).

4 Estimation strategy and results

Estimating the 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 of controlling for as many confounding factors as possible. While our results cannot be interpreted as definitely causal, we test their robustness in a series of subsequent analyses that convey a similar story as our baseline results and echo our discussions with academic founders. In what follows, we start by highlighting the determinants of founders' reliance on their academic work when founding their startup. We then present the OLS results on startup performance and proceed to the analysis of potential mechanisms behind our results. 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. Columns (1) and (2) estimate the association between the *Percentage self-cites* variable and founders' characteristics. Column (3) to (6) successively add the (log) number of publications published before startup creation, the (log) number of citations received before startup creation[11], the (log) amount of grant funding received before startup creation and the (log) number of unique co-authors before startup creation which we use as

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

[11]Citations are residualized in order to account for differences in publication year.

a proxy for network size. Overall, conditional on entering entrepreneurship and based on observable characteristics, professors who rely more on their academic work do not appear to be systematically different.

<Insert Table 2 here>

4.2 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, 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 has already founded one or several startups before.[12] We further control for the (log) average number of publications and the (log) average network size of founders before startup creation.[13] 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 main independent variable of interest. We use robust standard errors, clustered on the startup level.

Our dependent variables of interest focus on measures that proxy for 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., $U.S. 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.[14]

[12] 32 startups have what we call a "serial" founder.

[13] We average these variables at the startup level when there are several professor-founders. In practice, results are not significantly impacted by the inclusion of these variables.

[14] Table A10 adds sector × startup creation year fixed effects and finds similar results. Table A11 uses funding amounts from Thomson Reuters and finds similar results.

<Insert Table 3 here>

Column (1) displays that academic startups whose knowledge-base relies more on the scientific work of their founders (i.e., an increased percentage of self-cites) are associated with a lower probability of raising funding amounts that fall in the upper 25^{th} percentile. A 10p.p decrease in distance (i.e., a 10p.p increase in the percentage of self-cites) is associated with a 3.4% decrease in the probability of having raised more than $U.S. 10 million within 5 years of inception. Column (2) displays that academic startups whose knowledge-base relies more on the scientific work of their founders are associated with a lower likelihood of being acquired: a 10p.p decrease in distance (i.e., a 10p.p increase in the percentage of self-cites) is associated with a 2.2% decrease in acquisition likelihood. We do not find any significant effect for the probability of IPO.

5 Mechanisms

In this section, we investigate potential mechanisms driving our results. We successively discuss whether the lower success of startups whose knowledge-base is closer to their founders' academic work is primarily driven by (i) a lower technology quality, (ii) a more nascent technology, (iii) founders' role in the venture, or (iv) the nature of science they rely on.

5.1 Technology quality

In Table 4, 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 4 here>

Table 5 shows 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 raising subsequent funding amounts and the lower likelihood of acquisition experienced by startups whose knowledge-base is closer to their founders' academic work is not related to lower inventive output or worse innovation quality.

<Insert Table 5 here>

5.2 Stage 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 granted 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 the publication year from the patent application year and use the 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. Columns (2) and (4) of Table 6, which include all the controls 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 paper 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.[16]

<Insert Table 6 here>

[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.

[16]Hazard models, left unreported, also do not reveal differences in terms of "time to acquisition".

5.3 Founders' role

Founders who rely more on their own academic work may decide to keep a more hands-on role in their startup (Ding and Choi, 2011; Roche et al., 2020). This could result in poorer startup outcomes compared to scenarios where executive roles are assigned to individuals without a dual involvement in both Academia and entrepreneurship. We test this in several ways. First, we classify each founder depending on whether they have an active role in the company (e.g., Chief Executive Officer, Chairman, Chief Financial Officer) or not (e.g., an advising role for instance by sitting on the board). We run a founder-level regression of an indicator equal to 1 if the founder has an active role in the startup and 0 otherwise on *Percentage self-cites*. We use robust standard errors, clustered at the founder level. Results are presented in Table A1 and we do not find any statistically significant result.

We also explore whether founders who rely more on their own academic work experience worse outcomes on the academic side, which would be consistent with an escalation of commitment and challenges associated with managing a dual involvement. In order to assess the plausibility of this mechanism, we study the relationship between the distance in knowledge-bases and academic output. We implement a two-way fixed effects specification as follows:

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

where i indexes individuals and t indexes year. In our specification, $Rely_i$ equals 1 if individual i is part of the group relying on their science (that we define below) and 0 otherwise, while $Post_t$ equals 1 for years following startup creation. X_{it} comprises bins of experience dummies.[17] 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

[17] One dummy for years of experience 1-4, one dummy for years of experience 5-10, 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.

indicator with a 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. For both the two-way fixed effects and the event study specification, we use robust standard errors clustered at the founder level.

Our main dependent variables of interest are the number of publications, the number of top publications (defined as papers receiving more citations than the average number of citations of other papers in the same area of research and publication year) and the number of (year- and field-adjusted) citations. 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 369 founders with pre- and post-entrepreneurship information and 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 *Percentage self-cites* variable is above the mean[19], while the control group is composed of individuals whose *Percentage self-cites* variable is below the mean. We will explore the sensitivity of our results to other definitions of the treatment and control groups in subsequent analysis.

<Insert Table 7 here>

Table 7 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. We also do not find differences in founder-level characteristics. Panel B explores academic output at the time of startup creation. Both groups have, on average, slightly more than 20 years of experience when

[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

[19]The mean equals 7.9% in this sample

they began their entrepreneurial activity. There is no significant difference between the treatment and the control groups with regards to the 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 A1 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 up to startup creation. After that, it stabilizes and then seems to slightly decrease for the treatment group, while it 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 the research production function over a professor's lifetime.

We now go beyond the raw data and explore results in more detail in a regression framework. Figure A2 shows the β_τ coefficients of the event-study regression for the six outcomes of interest. While we cannot formally prove the parallel trend assumption necessary for a difference-in-difference analysis, it is reassuring that we find no apparent pre-trends. Table 8 displays 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 8 here>

Overall, these results cast doubt that founders' role and an escalation of commitment mechanism are the primary drivers behind our results.

5.4 Nature of Science

The lower success rate experienced by startups whose knowledge-base is closer to their founders' academic work could be explained by differences in the scientific knowledge used to create the technology. To explore this potential mechanism, we try to characterize the *nature* of science which startups build on. To that end, we analyze the papers which startups' first patent(s) build on. Each paper in the RoS dataset is associated with a Microsoft Academic Graph (MAG) identifier but we are not aware of any crosswalk between MAG and Dimensions AI. Hence, We use the Digital Object Identifier (DOI) and PubMed identifier (PMID) associated with each paper cited by the first patents of a startup to retrieve information about them using Dimensions AI. We focus on two main measures that capture (i) the *breadth* of the papers startups build on and (ii) the *type* of knowledge embodied, specifically, the extent to which the knowledge is specialized.

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 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 9 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 calculate the number of times it was used by other papers and average this measure at the startup level.[20] We conceptualize this measure as our proxy for the type of knowledge used, namely, 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 9 and show a significantly negative coefficient, implying that startups relying more on their founders' academic work build on scientific output 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 lower quality of the paper. Column (3) focuses on the citations that these papers received from patents (excluding patents from the focal startup).[21] Interestingly, startups closer to their founders' academic work also seem to build on scientific knowledge that tends to be used less by other patents. This is consistent with the fact that their more narrow and more specialized knowledge is potentially less easy to integrate into technological solutions.

<Insert Table 9 here>

[20]In order to account for differences in publication year, we standardize this measure following Perry and Reny (2016)

[21]This measure is also standardized in order to account for differences in publication year

6 Robustness

Goodman-Bacon (2021) provides 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 A2 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). We also explore the sensitivity of the difference-in-difference specification to the choice of the treatment and control groups. Table A3 shows the results when defining the treatment group as individuals with a strictly positive *Percentage self-cites* variable and the control group as individuals with a *Percentage self-cites* variable equal to 0. Again, we do not find any statistically significant result, confirming that an escalation of commitment is not likely to be the main mechanism behind the startup performance results.

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.[22] This leaves us with a sample of 185 startups, with a mean *Percentage of self-cites* variable of 13.5% (see Figure A3 in the Appendix for an histogram of the main independent variable of interest). 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 also replicate our results by using all patents pertaining to a startup to calculate the distance measure. In other words, 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. Table A7 shows similar results regarding startup performance, with a 10p.p increase in founders' reliance on their own academic work being associated with a 4.5% decrease in the probability of having raised more

[22]This excludes 110 startups that cite scientific papers but none of them come from the founders' previous academic work and 13 startups that do not cite any scientific paper.

than \$U.S. 10 million within 5 years of inception and a 1.7% lower likelihood of getting acquired, though the estimate on this last outcome becomes noisier. Again, we do not find any significant difference in innovation quality as proxied by number of patents (Table A8). We also do not find any significant difference with regards to technological development. Finally, we examine the papers cited by all the patents associated with each startup in order to characterize the science they use. Table A9 displays the results of this exercise. Here, again, we find that startups relying more on their founders' academic work build on science that is narrower and more specialized.

7 Discussion and Conclusion

Herein, we examine how the extent to which academic founders exploit their own technologically unique knowledge in the formation of their startup relates to entrepreneurial success after founding. To do so, we create an initial dataset comprising 510 academic startups corresponding to 676 founder-professors in bio-medicine. Our results regarding startup performance indicate that startups whose knowledge-base relies more on the previous academic work of their founders are less likely to be acquired and to raise a significant amount of funds. We provide evidence that these results do not seem to be linked to a lower invention quality, as proxied by the number of patents and the citations these patents receive. We also find no definitive support for the idea that startups closer to their founders' academic work are 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. Examining founders' involvement in their ventures as well as their outcomes on the academic market, differences in founders' role or an escalation of commitment do not seem to be the main drivers of our results.

Instead, we find indication that differences in the *nature* of the scientific knowledge employed may play a role. In particular, we find that startups that rely more on their founders' previous academic work tend to utilize knowledge that is more narrow and more specialized, despite being of similar 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 (Arora, Fosfuri and Roende, 2022; Bikard et al., 2019). It also aligns with our finding of having a lower propensity to raise significant amounts of funding, since VCs and other investors may face hurdles in evaluating the potential of highly specialized technologies. Furthermore, this mechanism echoes the absence of a detectable effect on IPO, since this outcome is more dependent on growth prospects once uncertainty around the technology has been resolved, and there is less need for integration with complementary assets (Bayar and Chemmanur, 2011).

As such, our findings highlight that the success of academic startups may be influenced not only by the incorporation of scientific knowledge, but also by the specific type and breadth of the knowledge that they use. These results underscore a crucial strategic balancing act academic founders may need to strike: leveraging specialized knowledge while considering its market applicability. Although using science in invention may be a source of advantage on the market for technologies (Arora, Belenzon and Suh, 2022), the *nature* of science being used also requires consideration since, as our results highlight, it plays an important role in achieving performance milestones.

In addition, 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. Intermediaries (e.g., Technology Transfer Offices) may be especially critical in this regard (`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 and continuously, objectively reassess if their original technology solves the problem it set out to solve. Beyond recognizing differences in institutional logics across academia and industry (Sauermann and Stephan, 2013), 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 targeted postdoctoral training 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 confounds 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 further deal with identification challenges. Future work may also examine 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

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, R., Kryscynski, D. and Singh, H. (2015), 'Evaluating venture technical competence in venture capitalist investment decisions', *Management Science* **61**(11), 2685–2706.

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.

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, Technical report, National Bureau of Economic Research.

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.

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

Bayar, O. and Chemmanur, T. J. (2011), 'IPOs versus acquisitions and the valuation premium puzzle: A theory of exit choice by entrepreneurs and venture capitalists', *Journal of Financial and Quantitative Analysis* **46**(6), 1755–1793.

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., 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.

Bryan, K. A., Ozcan, Y. and Sampat, B. (2020), 'In-text patent citations: A user's guide', *Research Policy* **49**(4), 103946.

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

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.

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

Ding, W. and Choi, E. (2011), 'Divergent paths to commercial science: A comparison of scientists' founding and advising activities', *Research policy* **40**(1), 69–80.
Geroski, P. A., Mata, J. and Portugal, P. (2010), 'Founding conditions and the survival of new firms', *Strategic Management Journal* **31**(5), 510–529.
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.
Kacperczyk, A. and Younkin, P. (2017), 'The paradox of breadth: The tension between experience and legitimacy in the transition to entrepreneurship', *Administrative Science Quarterly* **62**(4), 731–764.
Krieger, J. L., Schnitzer, M. and Watzinger, M. (2022), Standing on the shoulders of science, Technical report, Harvard Business School Working Paper.
Lee, H. (2023), 'The heterogeneous effects of patent scope on licensing propensity', *Research Policy* **52**(3), 104696.
Lin, J. and Wilbur, W. J. (2007), 'Pubmed related articles: A probabilistic topic-based model for content similarity', *BMC bioinformatics* **8**(1), 1–14.
Makri, M., Hitt, M. A. and Lane, P. J. (2010), 'Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions', *Strategic management journal* **31**(6), 602–628.
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.
Matthews, A. (2023), 'The idea and becoming of a university across time and space: Ivory tower, factory and network', *Postdigital Science and Education* **5**(3), 665–693.
Mowery, D. C., Nelson, R. R., Sampat, B. and Ziedonis, A. (2004), *Ivory Tower and Industrial Innovation*, Palo Alto: Stanford Business Books.
National Science Board (2018), *Science and Engineering Indicators 2016*, Arlington, VA: National Science Foundation.
Park, U. D., Rha, Y., Shah, S. K. and Chattopadhyay, S. (2023), 'An image of industry: exploring the effects of knowledge sources in the medical imaging industry', *Industrial and Corporate Change* **33**(1), 64–89.
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.
Perry, M. and Reny, P. J. (2016), 'How to count citations if you must', *American Economic Review* **106**(9), 2722–2741.
Polidoro, 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.
Roach, M. (2017), 'Encouraging entrepreneurship in university labs: Research activities, research outputs, and early doctorate careers', *PLOS ONE* **12**(2).
Roach, M. and Sauermann, H. (2015), 'Founder or joiner? the role of preferences and context in shaping different entrepreneurial interests', *Management Science* **61**(9), 2160–2184.
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.

Sauermann, H. and Stephan, P. (2013), 'Conflicting Logics? A Multidimensional View of Industrial and Academic Science', *Organization Science* **24**(3), 889–909.

Schmidt, J. B. and Calantone, R. J. (2002), 'Escalation of commitment during new product development', *Journal of the academy of marketing science* **30**, 103–118.

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. (1981), 'The escalation of commitment to a course of action', *Academy of management Review* **6**(4), 577–587.

Stenard, B. S. and Sauermann, H. (2016), 'Educational mismatch, work outcomes, and entry into entrepreneurship', *Organization Science* **27**(4), 801–824.

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

Stuart, T. E., Hoang, H. and Hybels, R. C. (1999), 'Interorganizational endorsements and the performance of entrepreneurial ventures', *Administrative science quarterly* **44**(2), 315–349.

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.

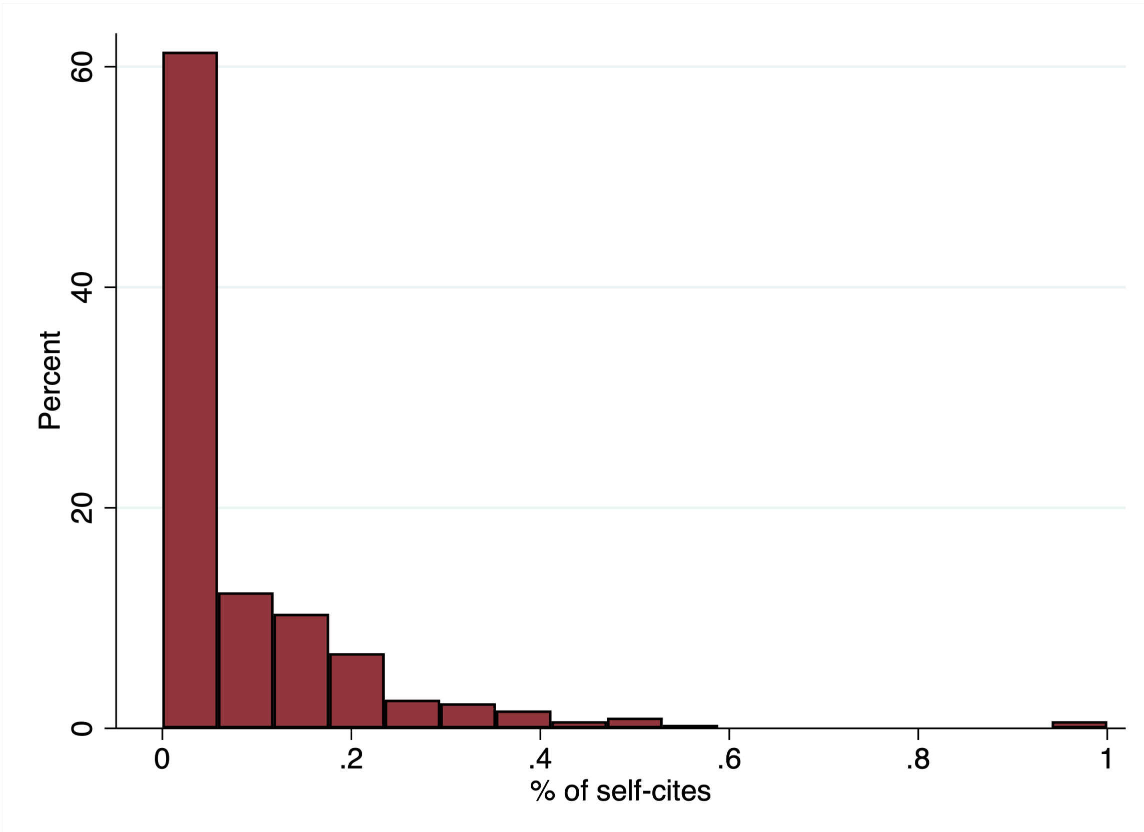
Wasserman, N. (2017), 'The throne vs. the kingdom: Founder control and value creation in startups', *Strategic Management Journal* **38**(2), 255–277.

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: Histogram of the main independent variable of interest



Notes: This figure displays the distribution of the *Percentage self-cites*, which represents the percentage of scientific citations in a startup's (first granted) patents that come from papers written by founders themselves.

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 citing science	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 displays summary statistics for our sample of 308 academic startups.

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

	Percentage self-cites					
	(1)	(2)	(3)	(4)	(5)	(6)
Experience=[5,10]	-0.00908	-0.00961	-0.0103	-0.0149	-0.0144	-0.0122
	(0.0408)	(0.0420)	(0.0423)	(0.0431)	(0.0432)	(0.0412)
Experience=[11,15]	-0.0293	-0.0257	-0.0268	-0.0310	-0.0303	-0.0284
	(0.0416)	(0.0431)	(0.0457)	(0.0463)	(0.0471)	(0.0452)
Experience=[16,20]	0.0101	0.00951	0.00815	0.00412	0.00474	0.00470
	(0.0436)	(0.0450)	(0.0469)	(0.0470)	(0.0485)	(0.0467)
Experience=[21,25]	-0.0218	-0.0224	-0.0240	-0.0263	-0.0256	-0.0242
	(0.0401)	(0.0415)	(0.0452)	(0.0458)	(0.0467)	(0.0448)
Experience$\geq$25	0.00373	0.00398	0.00198	0.000941	0.00133	-0.000301
	(0.0401)	(0.0415)	(0.0479)	(0.0485)	(0.0491)	(0.0475)
Top Institution		0.00211	0.00207	0.00283	0.00280	0.00321
		(0.0110)	(0.0110)	(0.0109)	(0.0110)	(0.0111)
Female		-0.0209	-0.0207	-0.0202	-0.0201	-0.0189
		(0.0142)	(0.0144)	(0.0142)	(0.0143)	(0.0142)
Serial founder		-0.0284	-0.0285	-0.0259	-0.0259	-0.0268
		(0.0186)	(0.0188)	(0.0188)	(0.0188)	(0.0186)
Publications$_{t-1}$			0.000579	0.0107	0.0111	0.0166
			(0.00677)	(0.0110)	(0.0121)	(0.0126)
Citations$_{t-1}$				-0.0116	-0.0115	-0.00820
				(0.0103)	(0.00986)	(0.0102)
Grants$_{t-1}$ (log)					-0.000186	-0.0000512
					(0.00145)	(0.00149)
Network size (log)						-0.0108
						(0.00909)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	443	443	443	443	443	443
R-sq.	0.163	0.169	0.169	0.172	0.172	0.174

Notes: This table displays 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 citations received before startup creation. Citations are adjusted for publication year by dividing their value by the average number of citations received by articles with the same publication year (Perry and Reny, 2016). $Grants_{t-1}$ is the total amount of grants received before startup creation. *Network size* is the number of unique co-authors before startup creation. We add state, sector and startup creation year fixed effects in each model. Standard errors (in parentheses) are clustered at the founder level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: OLS Performance Outcomes, Startup Level

	$\mathbf{1}\{\text{Funds} > p(75)\}$ (1)	Acquired (2)	IPO (3)
Percentage self-cites	-0.417**	-0.248***	0.109
	(0.175)	(0.0952)	(0.186)
Number patents (log)	-0.0328	-0.0424	0.0312
	(0.0783)	(0.0494)	(0.0765)
Number scient. patents (log)	0.0871	0.0471	0.0221
	(0.0746)	(0.0486)	(0.0690)
Team size (log)	0.0336	0.0606	-0.0331
	(0.0974)	(0.0538)	(0.0552)
At least one female founder	0.0146	-0.0997***	0.000753
	(0.0779)	(0.0336)	(0.0423)
At least one top-tier university	0.125**	0.0145	0.0214
	(0.0601)	(0.0350)	(0.0429)
At least one serial founder	-0.0801	-0.0194	0.0659
	(0.109)	(0.0650)	(0.0911)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	308	308	308
R-sq.	0.204	0.205	0.166

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 $US 10 million (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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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: OLS Innovation Outcomes, Startup Level

	Nb patents (log) (1)	Nb patents (Poisson) (2)
Percentage self-cites	-0.194	-0.126
	(0.337)	(0.523)
Team size (log)	-0.0216	-0.0326
	(0.168)	(0.236)
At least one female founder	-0.185	-0.288*
	(0.124)	(0.165)
At least one top-tier university	-0.0414	0.00647
	(0.110)	(0.165)
At least one serial founder	0.242	0.224
	(0.187)	(0.218)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	308	308
R-sq.	0.200	

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 (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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 5: Forward Citations, Patent Level

	Forward citations (Poisson)			
	First Patents		All Patents	
	(1)	(2)	(3)	(4)
Percentage self-cites	0.355	0.721	-0.466	-0.225
	(0.487)	(0.439)	(0.496)	(0.434)
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. Columns (2) and (4) add patent class fixed effects. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders) and we add patent application year and sector 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 6: OLS Stage of Development Outcomes, Startup Level

	Time from paper to patent		Time from paper to startup	
	(1)	(2)	(3)	(4)
Percentage self-cites	-14.22**	-11.17	-9.724	-5.116
	(6.889)	(7.138)	(6.527)	(6.744)
Team size (log)		2.171		2.793
		(2.856)		(3.029)
At least one female founder		-2.004		-0.367
		(2.417)		(2.472)
At least one top-tier university		-1.842		-1.933
		(2.054)		(2.098)
At least one serial founder		2.614		1.766
		(3.480)		(3.430)
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.	0.0227	0.178	0.0133	0.162

Notes: Each observation corresponds to a startup. In columns (1) and (2), the outcome variable is the number of years between the application year of a startup's first patent(s) and the publication year of its first scientific cite. In columns (3) and (4), the outcome is the number of years between a startup's creation year and the publication year of its first scientific cite. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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 7: Summary statistics for the treatment and control groups of the difference-in-difference analysis

	Control group	Treatment group	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	137	0.94
Top publications (cumulative)	51	49	0.71
Citations ('000s) (cumulative)	16.9	14.0	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	244	125	

Notes: The treatment (resp. control) group is composed of individuals whose *Percentage self-cites* variable is above (resp. below) the mean, equal to 7.9%. *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 8: Difference-in-Difference Analysis, Academic Output

	Publications (log) (1)	Top (log) (2)	Cites (log) (3)	% first (log) (4)	% last (log) (5)
$\text{Treated}_i \times \text{Post}_t$	-0.0261	-0.00217	-0.0193	0.0122	0.0123
	(0.0679)	(0.0512)	(0.0507)	(0.0103)	(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	0.631	0.430	0.346	0.317

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. The treatment (resp. control) group includes individuals with a *Percentage self-cites* variable above (resp. below) the mean. All outcomes are logged. In each model, we add year fixed effects, individual fixed effects and experience-bin fixed effects. Standard errors (in parentheses) are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

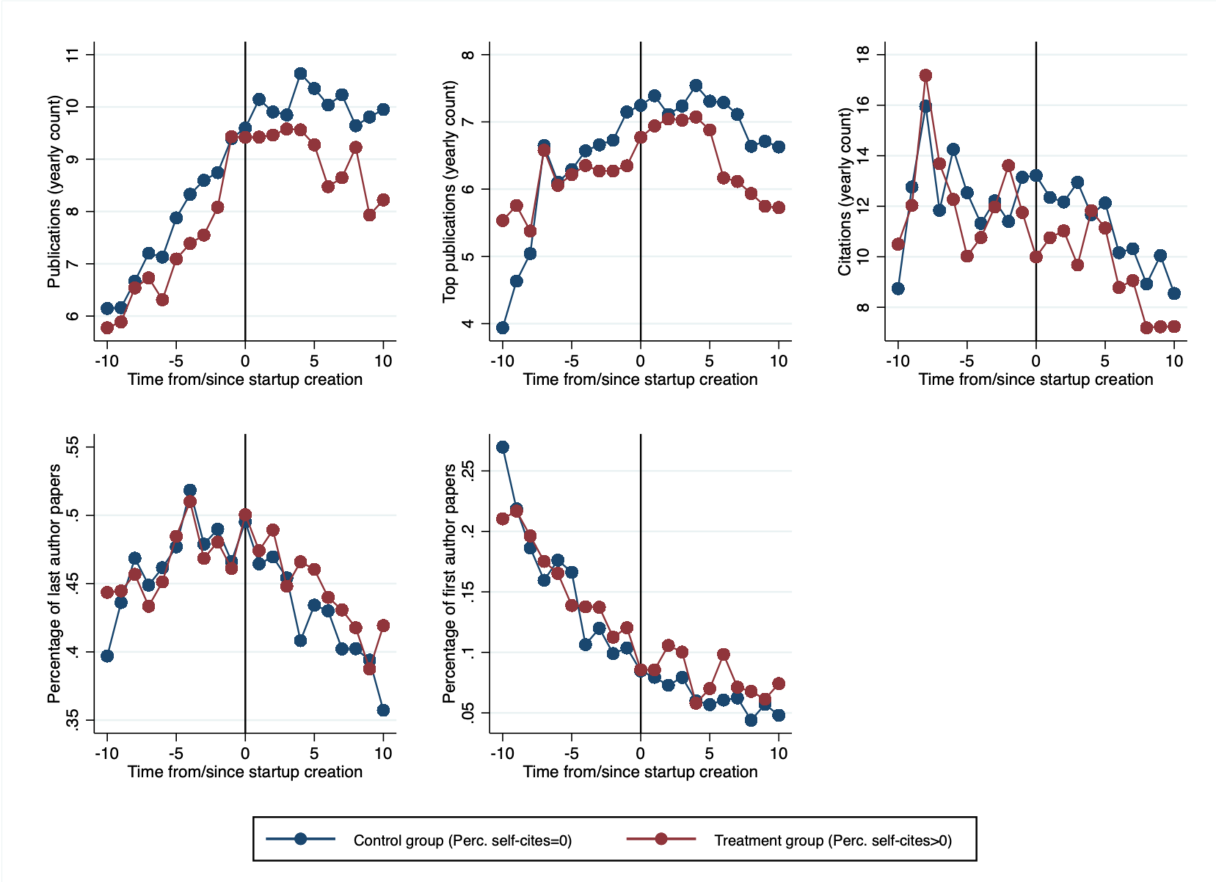
Table 9: Characteristics of the science cited by patents, Startup Level

	Concepts		Cites from patents
	Number (log) (1)	Use in other papers (log) (2)	(log) (3)
Percentage self-cites	-0.629**	-0.391***	-0.378*
	(0.272)	(0.137)	(0.216)
Number patents (log)	-0.0510	-0.175***	-0.0584
	(0.107)	(0.0578)	(0.107)
Number scient. patents (log)	0.0595	0.178***	0.0843
	(0.113)	(0.0593)	(0.110)
Team size (log)	0.0543	0.0289	0.0171
	(0.0600)	(0.0530)	(0.0881)
At least one female founder	-0.0190	0.0357	-0.183**
	(0.0494)	(0.0444)	(0.0712)
At least one top-tier university	-0.0286	0.00817	0.0219
	(0.0391)	(0.0359)	(0.0617)
At least one serial founder	-0.0610	0.0465	0.284**
	(0.0676)	(0.0577)	(0.116)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	283	283	288
R-sq.	0.197	0.247	0.246

Notes: Each observation corresponds to a startup. We focus on the papers cited by the first patents of a startup. In column (1), the outcome is the (log) average number of concepts used in these papers. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. This measure is standardized to account for differences in publication year. In column (3), the outcome is the number of citations that these papers received from patents (excluding patents from the focal startup). This measure is standardized to account for differences in publication year. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). We add sector, state, startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

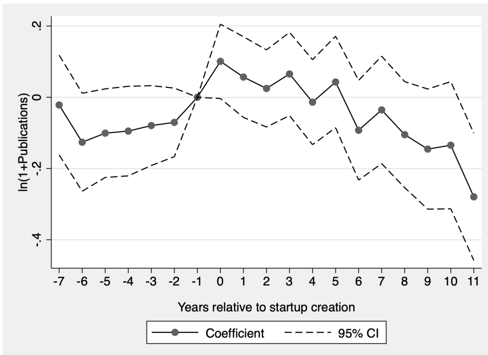
Figure A1: Raw plot of the academic outcomes for the treatment and control groups



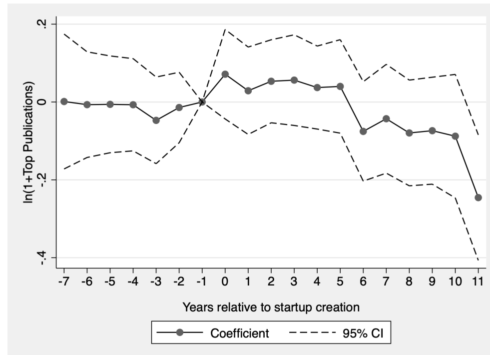
Notes: This figure shows the mean values before startup creation (negative values of the x-axis) and after startup creation (positive values of the x-axis) of several outcome variables for the treatment (red line) and control (blue line) groups. The treatment (resp. control) group includes individuals with a *Percentage self-cites* variable above (resp. below) the mean. 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 A2: Event study graphs, Academic outcomes

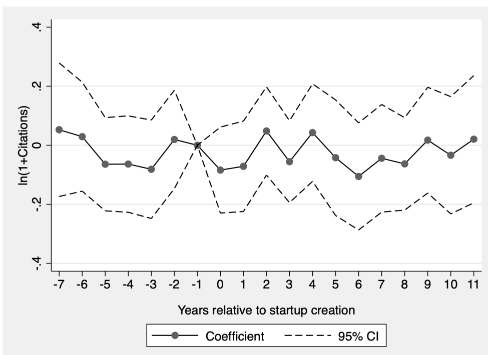
(a) Log(1+Publications)



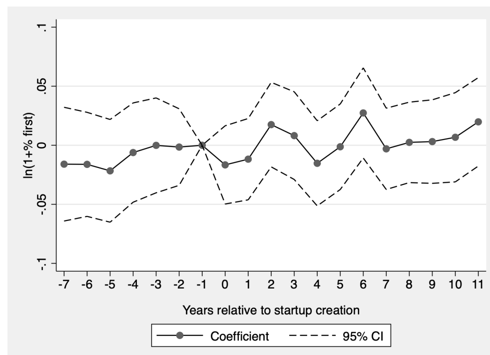
(b) Log(1+Top Publications)



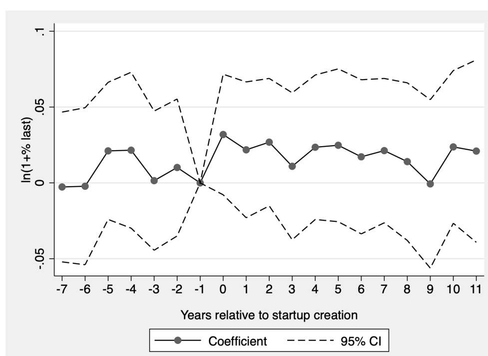
(c) Log(1+Citations)



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

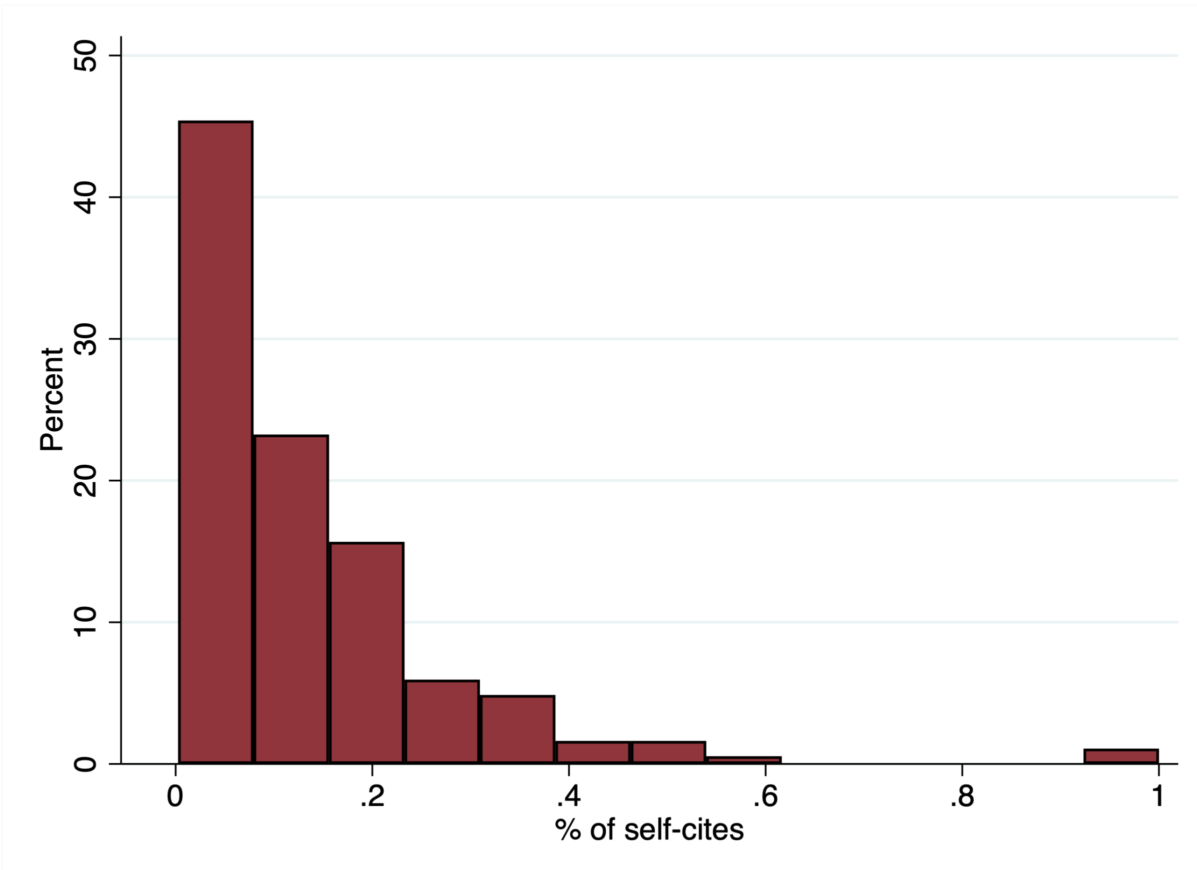


(e) Log(1+% papers as last author)



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 (resp. control) group is composed of individuals whose *Percentage self-cites* variable is above (resp. below) the mean. 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 A3: Histogram of the main independent variable of interest 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).

Table A1: Founders' Role

	Active Role		
	(1)	(2)	(3)
Percentage self-cites	0.0806	0.0413	0.0486
	(0.203)	(0.203)	(0.195)
Team size (log)	-0.533***	-0.524***	-0.496***
	(0.0585)	(0.0650)	(0.0651)
Female		-0.0471	-0.0446
		(0.0528)	(0.0535)
Top institution		0.0441	0.0507
		(0.0552)	(0.0544)
Serial founder		-0.212***	-0.172**
		(0.0724)	(0.0704)
$\text{Publications}_{t-1}$ (log)			-0.0133
			(0.0343)
Network size (log)			-0.0599
			(0.0410)
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	441	441	441
R-sq.	0.177	0.190	0.214

Notes: The dependent variable is an indicator equal to 1 if the founder has an active role in the startup and 0 otherwise. There are 407 unique founders but some of them are associated with more than one startup. The dependent variable is an indicator equal to 1 if the founder is part of the executive team of the venture and 0 otherwise. $\textit{Publications}_{t-1}$ is the aggregate number of publications published before startup creation. *Network size* is the number of unique co-authors before startup creation. We add sector and startup creation year fixed effects in each model. Standard errors (in parentheses) are clustered at the founder level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Difference-in-Difference Analysis, Academic Output - Goodman-Bacon

	Publications (log) (1)	Top (log) (2)	Cites (log) (3)	% first (log) (4)	% last (log) (5)
$\text{Treated}_i \times \text{Post}_t$	-0.0185	-0.0133	-0.0255	0.0108	0.0172
	(0.0685)	(0.0537)	(0.0522)	(0.0104)	(0.0154)
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes
Observations	82,955	43,108	43,108	74,273	74,273
R-sq.	0.551	0.644	0.426	0.355	0.323

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 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. The treatment (resp. control) group includes individuals with a *Percentage self-cites* variable above (resp. below) the mean. All outcomes are logged. In each model, we add year fixed effects, individual fixed effects 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, Academic output - Robustness of the choice of treatment and control groups

	Publications (log) (1)	Top (log) (2)	Cites (log) (3)	% first (log) (4)	% last (log) (5)
$\text{Treated}_i \times \text{Post}_t$	-0.0134	0.0317	-0.0148	0.00628	0.00292
	(0.0625)	(0.0414)	(0.0423)	(0.00899)	(0.0128)
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Experience-bin FE	Yes	Yes	Yes	Yes	Yes
Observations	13345	7091	7091	11986	11986
R-sq.	0.547	0.631	0.430	0.346	0.316

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. The treatment (resp. control) group includes individuals with a *Percentage self-cites* variable strictly positive (resp. equal to 0). 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

	$\mathbf{1}\{\text{Funds} > p(75)\}$ (1)	Acquired (2)	IPO (3)
Percentage self-cites	-0.401*	-0.219*	0.192
	(0.221)	(0.115)	(0.224)
Number patents (log)	-0.191**	-0.0589	0.0146
	(0.0923)	(0.0484)	(0.159)
Number scient. patents (log)	0.235**	0.0486	0.0596
	(0.103)	(0.0443)	(0.162)
Team size (log)	0.0138	0.0410	-0.0735
	(0.119)	(0.0635)	(0.0785)
At least one female founder	0.0636	-0.0966**	-0.0207
	(0.101)	(0.0385)	(0.0513)
At least one top-tier university	0.191**	0.0276	0.0304
	(0.0863)	(0.0416)	(0.0629)
At least one serial founder	0.0525	0.00102	0.160
	(0.155)	(0.0821)	(0.146)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	185	185	185
R-sq.	0.270	0.303	0.216

Notes: Each observation corresponds to a startup. We restrict observations to startups with *Percentage self-cites*>0. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than $US 10 million (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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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, Startup Level

	Nb patents (log) (1)	Nb patents (Poisson) (2)
Percentage self-cites	-0.420	-0.599
	(0.412)	(0.520)
Team size (log)	-0.00818	-0.0811
	(0.219)	(0.285)
At least one female founder	-0.137	-0.219
	(0.175)	(0.214)
At least one top-tier university	0.0398	0.168
	(0.158)	(0.199)
At least one serial founder	0.106	0.199
	(0.303)	(0.355)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	185	185
R-sq.	0.226	

Notes: Each observation corresponds to a startup. We restrict observations to startups with *Percentage self-cites*> 0. The outcome is the aggregate number of U.S. granted patents of a startup. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders), 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 patents, *Percentage self-cites*>0, Startup Level

	Concepts		Cites from patents
	Number (log)	Use in other papers (log)	(log)
	(1)	(2)	(3)
Percentage self-cites	-0.888***	-0.778***	-0.633**
	(0.338)	(0.143)	(0.268)
Number patents (log)	0.0147	-0.122***	0.0106
	(0.0601)	(0.0405)	(0.0830)
Number scient. patents (log)	-0.0215	0.137***	-0.00996
	(0.0638)	(0.0435)	(0.0851)
Team size (log)	0.0163	-0.00137	-0.0754
	(0.0683)	(0.0579)	(0.126)
At least one female founder	0.0237	0.0350	-0.0981
	(0.0494)	(0.0517)	(0.0956)
At least one top-tier university	-0.00232	-0.00949	0.0148
	(0.0463)	(0.0397)	(0.0876)
At least one serial founder	0.0109	-0.0412	0.217
	(0.0610)	(0.0633)	(0.172)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	181	181	182
R-sq.	0.325	0.406	0.301

Notes: Each observation corresponds to a startup. We restrict observations to startups with *Percentage self-cites*>0. We focus on the papers cited by the first patents of a startup. In column (1), the outcome is the (log) average number of concepts used in these papers. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. This measure is standardized to account for differences in publication year. In column (3), the outcome is the number of citations that these papers received from patents (excluding patents from the focal startup). This measure is standardized to account for differences in publication year. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). We add sector, state, 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 A7: OLS Performance Outcomes, Startup Level, All Patents

	$\mathbf{1}\{\text{Funds} > p(75)\}$	Acquired	IPO
	(1)	(2)	(3)
Percentage self-cites	-0.606***	-0.183	-0.0694
	(0.214)	(0.121)	(0.160)
Number patents (log)	-0.0380	-0.0358	0.0202
	(0.0785)	(0.0492)	(0.0769)
Number scient. patents (log)	0.0916	0.0408	0.0325
	(0.0747)	(0.0483)	(0.0698)
Team size (log)	0.0368	0.0662	-0.0387
	(0.0961)	(0.0540)	(0.0554)
At least one female founder	0.0151	-0.0971***	-0.00233
	(0.0773)	(0.0334)	(0.0420)
At least one top-tier university	0.120**	0.0136	0.0201
	(0.0598)	(0.0350)	(0.0427)
At least one serial founder	-0.0810	-0.0163	0.0615
	(0.107)	(0.0650)	(0.0911)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	308	308	308
R-sq.	0.204	0.205	0.166

Notes: Each observation corresponds to a startup. We use all the patents related to a startup to calculate *Percentage self-cites*. In column (1), the outcome is an indicator variable equal to 1 if the startup raised more than \$US 10 million (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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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, Startup Level, All Patents

	Nb patents (log) (1)	Nb patents (Poisson) (2)
Percentage self-cites	-0.273	-0.536
	(0.416)	(0.574)
Team size (log)	-0.0199	-0.0446
	(0.168)	(0.240)
At least one female founder	-0.185	-0.283*
	(0.124)	(0.166)
At least one top-tier university	-0.0436	-0.00163
	(0.110)	(0.165)
At least one serial founder	0.241	0.208
	(0.186)	(0.215)
State FE	Yes	Yes
Founding Year FE	Yes	Yes
Sector FE	Yes	Yes
Observations	308	308
R-sq.	0.200	

Notes: Each observation corresponds to a startup. We use all the patents related to a startup to calculate *Percentage self-cites*. The outcome is the aggregate number of U.S. granted patents of a startup. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders), 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: Characteristics of the science cited by patents, Startup Level, All Patents

	Concepts		Cites from patents
	Number (log)	Use in other papers (log)	(log)
	(1)	(2)	(3)
Percentage self-cites	-0.593**	-0.426***	-0.760***
	(0.246)	(0.149)	(0.225)
Number patents (log)	-0.0472	-0.174***	-0.0664
	(0.107)	(0.0577)	(0.108)
Number scient. patents (log)	0.0558	0.177***	0.0909
	(0.112)	(0.0592)	(0.111)
Team size (log)	0.0650	0.0344	0.0175
	(0.0636)	(0.0524)	(0.0890)
At least one female founder	-0.0147	0.0375	-0.186***
	(0.0493)	(0.0446)	(0.0706)
At least one top-tier university	-0.0324	0.00542	0.0170
	(0.0399)	(0.0362)	(0.0611)
At least one serial founder	-0.0617	0.0448	0.277**
	(0.0677)	(0.0574)	(0.116)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Observations	283	283	288
R-sq.	0.185	0.247	0.263

Notes: Each observation corresponds to a startup. We use all the patents related to a startup to calculate *Percentage self-cites*. The outcomes focus on the papers cited by all the patents of a startup. In column (1), the outcome is the (log) average number of concepts used in these papers. In column (2), the outcome is the (log) average number of times the concepts used in the papers have been used in other papers. This measure is standardized to account for differences in publication year. In column (3), the outcome is the number of citations that these papers received from patents (excluding patents from the focal startup). This measure is standardized to account for differences in publication year. In each model, we control for the (log) number of publications published before startup creation (averaged across founders) and the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). We add sector, state, 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 A10: OLS Performance Outcomes, Startup Level

	$\mathbf{1}\{\text{Funds} > p(75)\}$ (1)	Acquired (2)	IPO (3)
Percentage self-cites	-0.364*	-0.235**	0.0458
	(0.191)	(0.0969)	(0.168)
Number patents (log)	-0.00989	-0.0369	0.0355
	(0.0805)	(0.0486)	(0.0766)
Number scient. patents (log)	0.0654	0.0406	0.0189
	(0.0759)	(0.0474)	(0.0691)
Team size (log)	0.0599	0.0765	-0.0271
	(0.0986)	(0.0559)	(0.0574)
At least one female founder	0.0213	-0.101***	0.00442
	(0.0797)	(0.0344)	(0.0431)
At least one top-tier university	0.136**	0.0208	0.0315
	(0.0601)	(0.0364)	(0.0442)
At least one serial founder	-0.101	-0.0350	0.0590
	(0.108)	(0.0663)	(0.0939)
State FE	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes
Founding Year × Sector FE	Yes	Yes	Yes
Observations	308	308	308
R-sq.	0.241	0.220	0.184

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 $US 10 million (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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). We add sector, state, startup creation year and sector × 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 A11: OLS Performance Outcomes, Startup Level

	log(Funds) (1)	$\mathbf{1}\{\text{Funds} > p(75)\}$ (2)	Acquired (3)	IPO (4)
Percentage self-cites	-1.496***	-0.422**	-0.248***	0.109
	(0.577)	(0.192)	(0.0952)	(0.186)
Number patents (log)	0.161	0.0797	-0.0424	0.0312
	(0.324)	(0.101)	(0.0494)	(0.0765)
Number scient. patents (log)	0.193	0.0463	0.0471	0.0221
	(0.301)	(0.0956)	(0.0486)	(0.0690)
Team size (log)	0.377	0.112	0.0606	-0.0331
	(0.260)	(0.0830)	(0.0538)	(0.0552)
At least one female founder	-0.367*	-0.123*	-0.0997***	0.000753
	(0.216)	(0.0663)	(0.0336)	(0.0423)
At least one top-tier university	0.0185	-0.00343	0.0145	0.0214
	(0.170)	(0.0556)	(0.0350)	(0.0429)
At least one serial founder	0.434	0.158	-0.0194	0.0659
	(0.332)	(0.111)	(0.0650)	(0.0911)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	308	308	308	308
R-sq.	0.333	0.327	0.205	0.166

Notes: Each observation corresponds to a startup. We use funding amount from Thomson Reuters. In column (1), the outcome is the log amount of funds. In column (2), the outcome is an indicator variable equal to 1 if the startup raised more than $US 5.4 million (representing the 75^{th} percentile of the distribution). In column (3), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (4), 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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders). 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 A12: OLS Performance Outcomes, *Percentage self-cites* > 0, Startup Level

	log(Funds) (1)	$\mathbf{1}\{\text{Funds} > p(75)\}$ (2)	Acquired (3)	IPO (4)
Percentage self-cites	-1.883**	-0.529**	-0.219*	0.192
	(0.748)	(0.252)	(0.115)	(0.224)
Number patents (log)	-0.107	-0.00713	-0.0589	0.0146
	(0.572)	(0.156)	(0.0484)	(0.159)
Number scient. patents (log)	0.348	0.105	0.0486	0.0596
	(0.579)	(0.161)	(0.0443)	(0.162)
Team size (log)	0.365	0.143	0.0410	-0.0735
	(0.364)	(0.113)	(0.0635)	(0.0785)
At least one female founder	-0.526*	-0.147*	-0.0966**	-0.0207
	(0.289)	(0.0876)	(0.0385)	(0.0513)
At least one top-tier university	0.0352	-0.0141	0.0276	0.0304
	(0.257)	(0.0865)	(0.0416)	(0.0629)
At least one serial founder	0.159	0.0658	0.00102	0.160
	(0.519)	(0.161)	(0.0821)	(0.146)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	185	185	185	185
R-sq.	0.403	0.405	0.303	0.216

Notes: Each observation corresponds to a startup. We use funding amount from Thomson. We restrict observations to startups with *Percentage self-cites* > 0. In column (1), the outcome is the log amount of funds. In column (2), the outcome is an indicator variable equal to 1 if the startup raised more than $US 5.4 million (representing the 75^{th} percentile of the distribution). In column (3), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (4), 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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders), 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 A13: OLS Performance Outcomes, Startup Level, All Patents

	log(Funds) (1)	$\mathbf{1}\{\text{Funds} > p(75)\}$ (2)	Acquired (3)	IPO (4)
Percentage self-cites	-1.890***	-0.506**	-0.183	-0.0694
	(0.639)	(0.218)	(0.121)	(0.160)
Number patents (log)	0.158	0.0802	-0.0358	0.0202
	(0.329)	(0.103)	(0.0492)	(0.0769)
Number scient. patents (log)	0.195	0.0454	0.0408	0.0325
	(0.305)	(0.0968)	(0.0483)	(0.0698)
Team size (log)	0.394	0.117	0.0662	-0.0387
	(0.261)	(0.0832)	(0.0540)	(0.0554)
At least one female founder	-0.362*	-0.121*	-0.0971***	-0.00233
	(0.214)	(0.0657)	(0.0334)	(0.0420)
At least one top-tier university	0.00373	-0.00727	0.0136	0.0201
	(0.169)	(0.0551)	(0.0350)	(0.0427)
At least one serial founder	0.437	0.159	-0.0163	0.0615
	(0.332)	(0.111)	(0.0650)	(0.0911)
State FE	Yes	Yes	Yes	Yes
Founding Year FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Observations	308	308	308	308
R-sq.	0.334	0.326	0.198	0.164

Notes: Each observation corresponds to a startup. We use funding amount from Thomson. We use all the patents related to a startup to calculate *Percentage self-cites*. In column (1), the outcome is the log amount of funds. In column (2), the outcome is an indicator variable equal to 1 if the startup raised more than $US 5.4 million (representing the 75^{th} percentile of the distribution). In column (3), the outcome *Acquired* is an indicator variable equal to 1 if the startup is acquired. In column (4), 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 (log) number of publications published before startup creation (averaged across founders), the (log) of network size (proxied by the number of unique co-authors before startup creation and averaged across founders), we add sector, state and startup creation year fixed effects and cluster standard errors (in parentheses) at the startup level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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, Dimensions provides a disambiguation of researchers where a variety of variables are taken into account, 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 authors' profiles.[23] Our algorithm relies on identifying the 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 the 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 whom there were a high number of researcher IDs' matches (11 and 12 matches respectively)
- For researchers where no middle name is present 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-