When Do Ventures Founded by Research-oriented Entrepreneurs Outperform?*

Suting Hong, Kaihao Guo, Haipeng Zhang & David H. Hsu

August 2022

Abstract: Holders of a STEM doctoral degree are increasingly pursuing non-academic careers for a variety of reasons. There has been special interest in one such path, founding new ventures, because of its potentially disproportionate impact on societal and economic outcomes. We explore the conditions under which "research-oriented" founders (operationalized as those who have published in the peer-reviewed scientific literature) outperform otherwise observationally similar non-research-oriented founders in venture performance. Our contingency-based inquiry helps reduce tension in the literature which has found seemingly conflicting results on performance advantages to such academic founders. We also extend the human capital construct to the founding team and investor level, as such participants may assist in deficiencies in knowledge and expertise at the individual level. Our data consists of a matched sample of several thousand ventures across a wide span of industrial fields. Our results suggest that research-oriented founders' ventures outperform when they start ventures in their field of expertise, especially so when they are academically accomplished. Furthermore, for research-oriented founders starting ventures outside of their domain of expertise, recruiting co-founders with complementary functional backgrounds can compensate, though the same statement cannot be said about investors with complementary skills. We end by discussing policy implications.

Keywords: research-oriented founders; human capital; venture performance; venture capital.

* Suting Hong (email: hongst@shanghaitech.edu.cn); Kaihao Guo (emal: guokh@shanghaitech.edu.cn); Haipeng Zhang (email: zhanghp@shanghaitech.edu.cn); David H. Hsu (email: dhsu@wharton.upenn.edu). We thank participants of the 10th AIEA-NBER conference at KAIST for valuable comments. Hsu acknowledges funding from the Mack Center for Innovation Management at the University of Pennsylvania.

1. Introduction and motivation

The US and UK have produced many more STEM PhDs than can be accommodated in tenure-track academic positions (Alberts, et al. 2014; Blank, et al. 2017). The degree of imbalance differs by field, but several reports estimate that perhaps only some 10% of PhD recipients in STEM fields on average secure an academic post within 5 years of receiving their doctoral degree (Blank, et al. 2017). While there are a variety of institutional and preference-based reasons for this enduring imbalance[1], one implication is that there are many highly trained people who will seek employment in non-academic domains. Indeed, we are well past the point in the US in which more STEM PhDs enter industry as compared to going into academia (Langin, 2019), while acknowledging that individual sorting by preferences at least partly explains this pattern (Roach & Sauermann, 2010). According to the 2020 Survey of Earned Doctorates (the latest available as of this writing) administered by the US National Center for Science and Engineering Statistics[2], over 55,000 earned doctoral degrees were awarded in 2020 alone in the United States (which itself reflects a growing time trend). Using the rough metric of over half of these awardees entering non-academic employment (Langin, 2019), there are some 23,000 individuals annually with highly specialized training in this category.

One of the pathways within industry employment which is thought to be economically and societally important is technology commercialization, specifically via new venture development. A conventional wisdom is that "deep tech" entrepreneurs are disproportionately likely to be responsible for radical shifts in product development, resulting in extreme outcomes of enterprise value. While most of the prior work in this literature tends to focus on the spinout phenomenon, commercializing academic discoveries via startup formation (e.g., Marx & Hsu, 2022), our goal here is to examine the broader relationship between specialized human capital investments (as would be the case for STEM graduate degree holders) and venture outcomes. While we do not question the importance of academic spinouts, academic institutions likely have a broader impact in training graduate students who may found or join startups to commercialize technology (Roach

[1] Researchers in this space have offered a variety of explanations for persistent over-supply of graduate-level training. One explanation relates to incentives: graduate students are important in executing the scientific work in labs, and so principal investigators (PIs) have strong incentives to have a full pipeline of graduate researchers, even training many more individuals than would be necessary to replace the PIs themselves (Alberts, et al. 2014). The relatively stable (non-increasing) demand for growing most academic areas also contributes to the imbalance. In addition, other supply-side explanations suggest graduate degree holders' consumption-based preferences for conducting research and the high number of foreign trainees in the US (Alberts, et al. 2014).
[2] https://ncses.nsf.gov/pubs/nsf22300/data-tables (accessed May 3, 2022)

& Sauermann, 2015), even if such technology was not self-pioneered (as in the case of spinouts). Some prominent examples of such PhD-educated founders include the founders of Intel, Adobe, Snowflake, Databricks, and VMware, among many others.

Accordingly, we examine "research-oriented" founders, which we operationalize as having published in the peer-reviewed academic literature, as compared to otherwise observationally similar non-research-oriented founders. While this method of identifying specialized graduate training is imperfect (though academic publishing is most often in the domain of graduate-level skills and publishing is the "currency" of academia), it allows us to assemble a much larger sample (with a matched research-oriented and non-research-oriented founder sample of almost 6,000 spanning a range of industries) as compared to the prior literature. This is particularly important for this study, since we wish to examine contingency-based factors associated with venture outcomes (and so we require ample statistical power) as a way of reconciling opposing results reported in the existing literature (reviewed in section 2).

Several policy implications turn on the relationship between human capital and venture performance in this context, which makes the opposing results in the literature even more salient. One domain is in the area of graduate-level business and entrepreneurial skill training outside of the focal student's area of expertise, and whether/when such training improves venture outcomes. At a higher university policy and societal level, the extent of re-deployability of specialized (graduate-level) training and skill development relates to a more general notion of the individual ("consumptive") nature of such training as compared to a societal-level contribution (especially in recognition of the opportunity cost of such specialized training). While we do not attempt general social welfare analyses, our results may inform how university graduate training programs approach the decision of how many candidates to accept, especially given the considerable time and financial investments (including opportunity costs) of advisors, schools, and even governments. In addition, at the managerial policy level, the relationship direction between research-oriented founders and venture outcomes holds implications for investors, co-founders, and venture joiners from the perspective of best use of their financial and human resources.

Our data come from research-oriented founders identified through academic publication as detected through Microsoft Academic Graph (MAG) and the startups they found that raised external equity funding as identified through the Pitchbook database (the official data collector of the U.S. National Venture Capital Association). Our results suggest that research-oriented founders' ventures outperform when they start ventures in their field of expertise, especially so when they are academically accomplished. Furthermore, for research-oriented founders starting ventures outside of their domain of expertise, recruiting co-founders with complementary functional backgrounds can compensate, though the same statement cannot be said about investors with complementary skills.

2. Literature and Hypotheses

In this section we review the literature connecting human capital investments (particularly specialized investments, as is the circumstance of research-oriented founders) and entrepreneurial venture outcomes. Most of the literature contemplates this relation at the individual founder level, and so the first part of this section reviews that literature. Less discussed is the team-oriented nature of founding teams and their resource providers. We do so in the second part of this section. In both cases, our aim is to derive a set of empirically testable hypotheses specific to research-oriented founders.

2.1 Individual human capital and venture outcomes

We study research-oriented founders and conceptualize such individuals as having invested in a scholarly domain in sufficient depth to be capable of contributing to the peer- reviewed research literature. Defining individuals in this way blurs several distinctions the traditional literature makes, such as between university spinouts[3] (where the focal advance originated from the university and became the basis of a new business enterprise) and research-oriented training which may not have directly led to the focal start-up (e.g., Shah & Pahnke, 2014). The second group is likely to be a much larger one than the first, though the phenomenon of spinouts has received greater attention and the literature in that domain is much more developed (see, for example, the literature review on the subject contained in Rothaermel, et al. 2007). While we do not dispute the considerable challenges associated with entrepreneurial opportunity recognition, commitment, and venture resource development in the academic spinout process (Vohora et al.,

[3] The literature on academic spinouts/spinoffs is both wide ranging and important, though not squarely the focus of our study. To selectively discuss some facets of this literature, consider two interesting studies. Druilhe & Garnsey (2004) question the tendency in the literature to treat such ventures as homogenous, while Murray (2002) and others have explored different logics by which the same academic discovery can follow an academic paper pathway and a patenting pathway, as the former is the typical currency of academia while the latter is the typical currency of commerce and commercialization, and so patent-paper pairs can act as an intriguing (albeit specialized) window into development in each sphere of influence.

2004; Marx & Hsu, 2022), our aim here is to widen our investigative lens to examine broader (contingent) relations between research-oriented founders and venture success.

Human capital is often conceptualized as a factor of venture production, and an asset requiring investment (e.g., Becker, 1962; Coleman, 1988).[4] Accordingly, a substantial literature has developed examining general and specific educational training and work experience as important contributors to individual human capital. The common conceptualization across these two broad channels is that accumulated stocks of human capital can potentially confer returns in the future (invested skills may not immediately apply to the direct experience and skills necessary to be a successful entrepreneur). That general proposition is supported in the literature centered on the entrepreneurial context (see the literature review and meta-analysis in Unger, et al. 2011 based on 70 independent, published samples).

2.1.1 Educational training and entrepreneurship. Skill development through educational training can be of the general sort (often operationalized as years of education in the literature) or a specific type such as business and/or entrepreneurial training. In the former category, one early study connecting founders' individual human capital characteristics such as years of schooling to organizational survival found positive effects (Brüderl, et al. 1992). That finding has broad support in several other studies in the entrepreneurship literature (Unger et al., 2011), which draw from a wide range of empirical contexts and samples. In the case of research-oriented founders, the likely investments in their graduate school training transcends simple mastery of the corpus of existing knowledge in a domain. Instead, to publish in peer-reviewed academic journals, especially the most academically prestigious venues, a novel and creative process is also often required (not unlike the process of successful venture development).[5] In other words, a simple linear count of years of education (as is commonplace in the empirical human capital literature) may not suffice

[4] There are complex inter-relationships between various forms of organizational capital, which makes tracing "seed" conditions difficult. For example, studies have identified human capital investments as contributing to organizational social and financial capital (Hsu, 2007 reports that serial entrepreneurial founders are more likely to recruit executive officers for their ventures from their own social network, with implications for their ventures' financial valuation from venture capitalists, VCs). The literature has also reported that social capital in turn, can augment human capital (Coleman, 1988), thereby potentially leading to a self-reinforcing cycle. In addition, since human capital can both directly contribute to startup success and help attract VC investment (e.g., Colombo & Grilli, 2010), there are multiple channels in which human capital can operate in shaping venture outcomes. Our purpose here is not to disentangle these forms of individual and organizational capital, however.

[5] There are other parallels between STEM graduate training and the venturing process: both contexts require perseverance in the face of an unknown "recipe" for success, adhering to the scientific method for progress, and sometimes substantial lengths of time between evaluation events.

in predicting contributions to the scientific literature. This view is not lost in the literature on scientific publishing and careers (e.g., Levin & Stephan, 1991).

The other branch of the training literature may be conceptualized as more targeted training investments in business and/or entrepreneurship education. One of the large challenges in this literature is interpreting estimated effects as selection versus treatment effects, in that individuals do not randomly pursue such training. Often, those expecting to enter business or entrepreneurial contexts will be the same individuals who are motivated to seek out such training opportunities. A review of the entrepreneurial training literature, only some of which addresses this interpretational challenge, is contained in Martin et al. (2013). Even the most rigorous of the studies in this literature, those which utilize a field- or natural-experimental design, however, find results in opposing directions. On the one hand, Oosterbeek et al. (2010) report that entrepreneurship education does not map well to self-assessed entrepreneurial skills or intentions to enter entrepreneurship. Similarly, Karlan & Valdivia (2011), examining a developing market microfinance context, find little or no effect of a randomized training program on business revenues, profits, or employment. On the other hand, in a developed market context, also utilizing a randomized control trial design of entrepreneurial training, Camuffo et al. (2020) find evidence that such training enhances entrepreneurial decision-making and revenue. On balance, in a review and meta-analysis of 42 independent studies and samples, Martin et al. (2013) find support for a positive entrepreneurial training and education effect.

2.1.2 Work experience and entrepreneurship. Another branch of the human capital literature examines the role of individual experience. One piece of evidence on the importance of such experience is Azoulay et al.'s (2020) finding that the average age of successful founders in the U.S. is 45 years old. This suggests that the image of the college dropout being the modal successful entrepreneur may not be appropriate; rather, more seasoned and experienced individuals are more often associated with venture success.[6]

One set of studies examine a specific type of prior work experience: prior startup founding experience. While Cassar (2014) finds that such startup experience does not enhance entrepreneurs' focal venture performance forecasting, prior industry work experience does,

[6] The same issue about selecting into entrepreneurship based on existing human capital also applies to this discussion, in that the opportunity cost for highly experienced individuals may be higher than for those less experienced, and the level of experience may shape career choice (including engaging in new venture development).

especially in high-technology settings. Hsu (2007), on the other hand, finds evidence that prior founding experience facilitates hiring executive officers from the founders' own social networks, which in turn has financial implications for their ventures' valuations.

More generally, what is learned in work experiences which may contribute to subsequent venture success? In addition to being more familiar with customer needs and pain points (Shane, 2000), those with industry experience may be more likely to have the social contacts upstream (such as supplier networks) or downstream (relationships with distributors) to successfully execute their venture plan or gain information about product market opportunities as a result of such relations (Tripsas, 1997). They are also more likely to be aware of the competitive environment and the landscape of existing competitors in the market, and to pursue innovations industry incumbents do not pursue (Christensen, 1993; Klepper, 2001).

Related to this last set of studies, Helfat & Lieberman (2002) and others in the literature distinguish entrepreneurial spinoffs as having founders who were previously employed by an industry incumbent in the focal industry. Founders of such organizations may be unequal with respect to pre-entry resources. Helfat & Lieberman (2002), in reviewing the spinoff literature, suggest a perspective that these individuals inherit organizational resources, routines, and know-how which are somewhat broader in concept than the classic experience-based individual skill development. They find the empirical regularity that the greater the similarity of pre-entry resources and those needed to succeed, the more successful is the company. Finally, research-oriented founders may be more familiar with how to exploit a new technology based on prior scientific or technical training (Roberts, 1991).

The literature more narrowly on research-oriented founders and venture outcomes is limited, and points to a seemingly inconsistent set of results. While Colombo & Piva (2012) in a modest sample find evidence that such founders tend to invest and develop their businesses differently than non-research-oriented founders, Roche et al. (2020) examine their venture outcomes more directly. Using a sample of research-oriented founders in the biomedicine industries drawn from the Crunchbase dataset, Roche et al. (2020) find that such founders performance is worse than non-research-oriented founders (even while attaining similar external funding and patenting levels).

The results of this last study appear to be at odds with the resource advantages the literature has discussed. One reconciliation could be the contingent nature of research-oriented founders' human capital. Building on Helfat & Liberman's (2002) conclusion that the greater is the similarity and overlap between entrepreneurs' pre-entry resources and those needed for organizational success, we propose the following hypothesis:

- ***H1: Research-oriented founders will experience better venture outcomes as compared to observationally similar non-research-oriented founders when founding ventures in the domain of his/her research expertise.***

This prediction could hold either because research-oriented founders have a comparative advantage in identifying promising commercial applications of technical advances in their area of expertise, and/or because there is a perception (even if untrue) by holders of financial and human capital of such a research-oriented founder comparative advantage.

Furthermore, in line with Druilhe & Garnsey (2004) who argue that spinouts are not homogenous, we should also not think that research-oriented founders are a monolithic group. There is significant heterogeneity in such founders' degree of technical expertise and research quality (especially given the typical skewed distribution of research output and impact in academia). Roche, et al. (2020), for example, finds that superstar academic founders experience better venture outcomes relative to non-superstars. We therefore propose:

- ***H2: Among research-oriented founders, those with higher research productivity and quality will be associated with more favorable venture outcomes.***

2.2 Extended "team" complementary human capital and venture performance

If the above hypotheses are validated, then a salient question is recourse for research-oriented founders who are founding ventures *outside* of their focal expertise. There are two possibilities suggested by the literature, which center on the observation that founders are not atomistic individuals who are single-handedly charged with successfully launching their ventures. While much of the literature on human capital does not examine the team level, such a level seems appropriate. Examining team composition allows both specialization and breadth in a way which would be hard to replicate at the individual founder level. Indeed, across many areas of creative production, there has been a steady secular trend toward team production (Wuchty et al., 2007).

Functional diversity at the founding team level, while potentially attractive in achieving both breadth and depth, may not be easily attained, however. This is because of homophilic forces which typically govern social relationships, including forming entrepreneurial founding teams

(Ruef, et al. 2003). However, if research-oriented founders starting ventures outside of their domain of expertise (typically STEM) can recruit co-founders with complementary skills (such as those with business, marketing, and/or legal expertise), then venture performance might be enhanced (note: we take team composition as fixed in our empirical analysis; future efforts would ideally endogenize team composition).

Furthermore, an extended team structure might include not just cofounders, but also investors in the venture. Early-stage startup investors have been characterized as both "scouts" and "coaches". The "scout" role involves selecting the most promising venture ideas and entrepreneurs, while the "coach" role is aiding venture development via advisory roles and value-added services. If the second function is salient, as the literature suggests (e.g., Hsu, 2004), then the skills and expertise of the early-stage investor can serve as an extension of that of the founding team. We therefore have the following additional predictions:

- ***H3a: For research-oriented STEM founders starting ventures outside of their focal domain of expertise, assembling a founding team with complementary skills (e.g., business, marketing & law) will positively shape venture outcomes.***
- ***H3b: For research-oriented STEM founders starting ventures outside of their focal domain of expertise, having early-stage investors with complementary skills will positively shape venture outcomes.***

3. Data and Sample Construction

3.1 Data sources

We build our sample from the following sources: (i) data on investments and start-ups from Pitchbook; (ii) data on exit outcomes of funded start-ups from Pitchbook and Crunchbase; and (iii) data on individual researchers' publications from Microsoft Academic Graph.

Pitchbook provides comprehensive coverage of external equity investments in start-ups worldwide since 2007 to date. The data not only includes detailed information on investment rounds and funded companies, but also on individuals' educational and career records for founding team members[7] that are affiliated with those funded start-ups. The original data contain 268,805 founders affiliated with 164,742 start-up companies that received 339,580 investment rounds from 2007 to 2020 worldwide.

[7] Enterprise co-founders are labelled in the data, and at the time of first investment, the typical founding team size is small, with a mean of 2.4 and a median of 2 in our data.

The Microsoft Academic Graph (MAG) initiative constructs a heterogeneous graph consisting of scientific publication and citation relationships, as described in Wang et al. (2019) and Sinha et al. (2015). We download a snapshot of the entire MAG data through its authorized API in December 2020. The snapshot contains 245,252,418 records of publications or patents since 1800 to the date of downloading by 260,463,429 authors affiliated with 25,805 distinct institutions.

From Crunchbase, we extract information of exit events between January 2007 and November 2021 experienced by start-up companies that received external equity financing worldwide. We verify details of exit events (i.e., acquisition price, IPO status, etc.) from Crunchbase with records of exit events attained from Pitchbook.

3.2 Sample construction

We build a sample consisting of start-ups worldwide that received their first external equity investment round between the years 2007 and 2015.[8] The data provides detailed information about the investments, including size, valuation, participating investors, and exit status of start-ups. Furthermore, the data includes information about the educational and professional experience of individual members on the founding teams. Furthermore, applying data mining techniques from computer science, we develop an algorithm to detect whether any member of the founding team of those start-ups has published a research article in MAG dated prior to the first round of investment. For founders who have been successfully linked to authors in MAG by our criteria (as explained in Section 3.3.2), we then extract information on all their publications from MAG. Consistent with our theoretical frame, our empirical analysis uses a sample of research-founders that have published in the STEM fields.[9]

The remainder of this section is organized as follows. We explain in Section 3.2.1 how founders are linked to their author profiles in MAG. Section 3.2.3 describes how we assess whether a research-oriented founder launches their ventures in the same domain of their expertise. Section 3.2.3 introduces the method to match ventures with research-oriented founders to observably similar ventures founded by non-research-oriented individuals.

---

[8] Those first-round investments fall into the following four groups of deal types as denoted by Pitchbook: Seed Round (30%), Early-Stage VC (45%), Accelerator/Incubator (21%), and Angel (4%).
[9] In our sample, the majority (93%) of research-oriented founders have publications in the STEM fields. Nevertheless, all our results are robust to using the entire sample of research-oriented founders from all fields (STEM and non-STEM).

3.2.1 Linking founders in Pitchbook to authors in MAG. To link founders in Pitchbook to authors in MAG, we carry out two main tasks. *Step 1 is author name disambiguation in MAG*. MAG uses machine learning techniques to identify and disambiguate authors of scholarly publications (Wang et al. 2020). Built on what the MAG team has already done on author name disambiguation, we further conduct a comprehensive disambiguation following the approach by Sinatra et al. (2016). We explain the details of the algorithm in Appendix B.

Our *Step 2 is to match fields of publications in MAG to educational majors/concentrations in Pitchbook*. The similarity between the two is one of the key criteria for linking founders to their author profiles in MAG. We match educational majors/concentrations in Pitchbook to level-0 fields in MAG.[10] In total, there are 19 distinct level-0 fields covering major academic fields in MAG and 10,240 distinct entries of educational majors/concentrations in Pitchbook. For a given major/concentration, we find its sufficiently similar level-0 MAG field(s) by calculating its semantic similarities with all MAG fields of all levels (see Appendix B for a detailed explanation of the calculation procedure). In the end, we link 9,025 distinct entries of educational fields in Pitchbook to at least one level-0 field in MAG. Note that for a major involving inter-disciplinary knowledge, more than one level-0 field may be matched.

We then match founders in Pitchbook to their author profiles in MAG. We consider all the founding team members of start-ups worldwide for whom Pitchbook provides educational background information (i.e., educational institutions, fields of study, and year of graduation). We match founders in Pitchbook to the disambiguated author profiles in MAG by considering names, affiliations, field(s) of study, and first publication date. We discuss each in turn.

For each founder in Pitchbook, we locate all the authors in MAG who have the identical last names and first names.[11] For each founder in Pitchbook, we then compile a list of candidate authors in MAG as potential matches. We keep authors in the list of potential matches who have at least one affiliation in MAG and are like *either* educational institution affiliation *or* employment

[10] MAG use a six-level hierarchy in their taxonomy of fields of publications (assembled via a semantic understanding algorithm). There are a total of 694,121 distinct fields in the six-level hierarchy in MAG. However, *not all* publications are assigned to a field of study at *each* level of the hierarchy, and the broadest level (i.e., level 0) of fields shows the highest coverage rate (i.e., 98.78% as shown in Table A.1 Panel A in the Appendix). In total, there are 19 distinct level-0 fields covering major academic fields, and Table A.1 Panel B tabulates the counts of publications belonging to each level-0 field in the data.

[11] If only the initial letter of the first name is provided for a founder in Pitchbook, we look for all the authors in MAG with a first name of the same initial and the same last name.

position of a focal founder in Pitchbook.[12] Adopting the methods of Sinatra et al. (2016) and Levin et al. (2012), we compute the term frequency-inverse document frequency (TF-IDF) for words contained in the name of an affiliation. The TF-IDF is used to decide the weight of a word as we construct vector representation for the name of an affiliation.[13] We further decide name similarities based on cosine similarity of vector representations of affiliations, with a threshold of 0.8.

A matched author should have published in a field that is sufficiently similar to the education major/concentration of the focal founder. This happens when the level-0 field of at least one publication in MAG for an author is sufficiently similar with the major/concentration of a degree held by the founder, according to the procedure explained in *Step 2* above.

Finally, we require that the year of the first publication of a matched author should not be earlier than the year the person starts college, which is inferred by subtracting four years from the graduation year of his/her undergraduate degree as recorded in Pitchbook.[14]

3.2.2 Assembled sample. By considering all start-ups receiving external equity funding worldwide between 2007-2015, we link 8,060 founders (affiliated with 7,844 start-up companies) in Pitchbook to authors in MAG with publication records in the STEM fields.[15] We identify STEM fields according to the level-0 fields of publications as recorded in MAG. To assess the accuracy of the matching procedure described above, we randomly choose 200 matched pairs of authors in MAG and founders in Pitchbook. For each pair, we manually verify whether the author and the corresponding founder are the same person. We first gather authors' information from their homepages or Google Scholar profiles. We then collect information about entrepreneurs by searching company websites and LinkedIn profiles. Among the 200 pairs of authors and founders, we find the false positive rate to be 2% (four pairs are judged to be incorrect matches).

[12] We note that both Pitchbook and MAG can assign multiple institutions/organizations to a single person, and thus, we identify a potential match for any pair of author-founder if they share at least one sufficiently similar affiliated organization.

[13] As a commonly used method in computer science, TF-IDF measures the importance of a word to an individual document. In our setting, if a word appears in high frequency in the name of an affiliation but is rarely detected across the entire collection of affiliation names, the word will have a high TF-IDF score to the focal affiliation.

[14] Our results are robust to relaxing this restriction to allow the first publication to be no earlier than five years before the undergraduate graduation year.

[15] Among these founders, 38 distinct individuals are matched with more than one author. In those cases, we include in the sample the author with the highest number of papers among the potential matches.

After removing start-ups with missing covariate observations used in the empirical analysis (see Section 4 for details)[16], we have 3,836 start-ups headquartered in 59 countries/regions[17] and founded by individuals who hold publication records of 128,877 articles in the STEM fields in MAG. Table 1 Panel A shows a detailed tabulation of their publications by level-0 fields in MAG with most publications produced in the areas of Computer Science (22%) and Biology (16.7%).

3.2.3 Venture industry fields and domains of research expertise of research-oriented founders. To classify whether a research-oriented founder starts a venture in her domain(s) of research expertise, we examine the fields of publications authored by research-oriented founders and then compare to her venture's industry sector.[18] We use the following steps. First, Pitchbook uses an industry classification system comparable to the Global Industrial Classification Standard (GICS) that contains a hierarchy of industry classifications and industry sectors which at the broadest level contains seven distinct sectors.[19] Second, for each of those industry sectors of ventures, we identify level-0 publication fields in MAG that involve knowledge/expertise relevant to the focal sector. Appendix C provides a detailed description of how we link industry sectors of entrepreneurial firms to publication fields in MAG. If a founder has published *at least one* article in a field that is matched to the industry sector of her founded venture, we identify the founded venture is in the same domain of her research expertise.

We report in Table 1 Panel B the tabulation of industry sectors of 3,836 ventures launched by research-oriented founders in our sample. Columns 1 and 2 present distributions of those ventures across different industry sectors with the majority in R&D-intensive fields (i.e., 40% in Information Technology and 33% in Healthcare). Furthermore, in 56% of those ventures, the research-oriented founders' expertise is in line with the industry sector of their focal venture (as reported in Columns 3 and 4). When research-oriented founders launch a venture in an industry sector different from the domain of their research expertise (Columns 5 and 6), an increased share

[16] Most of the observation loss is caused by missing investor identity (17%) or deal investment size (16%). The two variables are used in our empirical estimation as described in Section 4.
[17] In Table A.2 in the Appendix, we tabulate the country locations of those ventures with the most headquartered in the United States (66%), followed by United Kingdom (7%) and Canada (3.7%).
[18] Alternatively, we rely on educational majors/concentrations rather than field codes from MAG publication records to identify domains of research expertise and re-run all our empirical analyses. All our results remain intact in both statistical and economic magnitudes.
[19] The seven industry sectors are "Information Technology", "Healthcare", "Materials and Resources", "Energy", "Business Products and Services (B2B)", "Consumer Products and Services (B2C)", and "Financial Services."

of ventures are built in fields that rely less on R&D and more on commercial development (i.e., 22% in B2B and 26% in B2C).

3.2.4 Building the control sample. To mitigate the role of unobserved differences between the founders, we build a matched sample of ventures comparable to those founded by research-oriented founders (but were founded by non-research-oriented individuals), using the following procedure. First, we start from ventures that were founded by research-oriented entrepreneurs and received their first external equity investments between 2007 and 2015 worldwide. We perform exact matching with ventures founded by non-research-oriented entrepreneurs based on the following variables: country location, industry (41 groups, as explained below), year of investment, and year of company foundation. We find a match for 2,770 ventures with research-oriented founders. Therefore, our sample contains 5,540 ventures worldwide receiving first-round equity investments between 2007 and 2015. Panel A Table 2 reports summary statistics of the sample that contains matched pairs of ventures with and without research-oriented founders.

4. Empirical results

4.1 Research-oriented founders versus non-research-oriented founders

In H1, we ask whether research-oriented founders experience better venture outcomes when developing enterprises within their domain of research expertise as compared to non-research-oriented founders. We track outcomes of ventures by examining their exit status by the time of data extraction (November 2021). We consider two alternative measures of venture exit outcome: *Success* and *Good Exit*. *Success* takes the value of one if an entrepreneurial firm achieved liquidity through an IPO or an M&A exit, and zero otherwise. The prior literature (e.g., Roche et al. 2020; Hochberg et al. 2007) uses similar measures to assess success of start-up companies. We also construct a more stringent measure for exit success, *Good Exit,* that equals one if a venture eventually experiences a good exit and zero otherwise. Following the literature (Ewens et al. (2016), Ewens and Marx (2018)), we identify a good exit as either an IPO or an acquisition with a known valuation that is at least two times the capital invested in the focal venture. In our empirical analysis, we estimate a linear probability model (LPM). We report results from LPM in the paper as we include a great variety of fixed effects in our specification (as explained below) which occasionally leads to convergence problems in non-linear estimation.[20] Nevertheless, our results remain intact using non-linear estimation methods.[21]

We estimate the following specification using a sample of matched pairs of ventures with and without research-oriented founders.

$$Y_{evct} = \beta_0 + \beta_1 ResearchFounder_e + \beta_2 C_{et} + \beta_3 X_{vt} + \phi_c + \tau_t + \tau_e + \epsilon_{evct} \quad (1)$$

where e, v, c, and t index the start-up company, lead investor, country, and investment year, respectively. The variable of interest, *ResearchFounder,* is binary, and equals to one if at least one founding team member of a focal start-up is research-oriented and zero otherwise. Y_{evct} denotes the dependent variable, namely *Success* or *Good Exit.*

C_{et} represents a set of controls for start-up company characteristics. Regarding team composition, we control for the logged number of founding team members (given the skewed distribution) at the time of the investment, *Ln (Team size)*, and a binary variable indicating whether there is any female founder on the team, *Female founder*. Furthermore, we include, *Serial founder,* that equals to one if at least one founding team member was on the founding team of a different venture that previously received external equity funding. Finally, we control for industry classification of entrepreneurial firms. Pitchbook uses an industry classification system comparable to the Global Industrial Classification Standard (GICS), in which entrepreneurial firms are grouped into 41 distinct industry groups. We construct dummies indicating the primary industry group that a venture is associated with.

X_{vt} is a set of controls for investor- and focal investment round-characteristics, including (a) logged number of investors in a focal round, (b) logged investment deal size, (c) deal type, and (d) logged experience level of the lead investor. The lead investor plays a vital role in the consummation of a deal by "providing an anchor investment, setting the valuation and instilling confidence in other potential investors based on their due diligence" (Pitchbook, 2020).[22] We quantify the experience of a lead investor using the number of prior investment rounds that a lead

[20] Wooldridge (2010, p. 563) suggests that "LPM often does a very good job if the main purpose of estimating a binary response model is to approximate the partial effects of the explanatory variables."

[21] See Table A.3 in the Appendix for results from estimating equation 1 via Probit models.

[22] The identity of lead investors is available for about 70% of the funding rounds in the entire sample. In the 30% of rounds where a flag for the lead investor is missing, we follow Ewens et al. (2022) and assume the lead investor is the investor with the largest number of years since its first investment at the time of the funding round. Unlike the Thomson One database, Pitchbook does not provide information on investment amount contributed by each individual investor. This constrains us from following the previous literature and relying on the per-firm investment amount to define lead investors.

investor participated in, *Lead Exp*. Furthermore, there are four different types of deals for the first investments in the sample: "Accelerator/Incubator", "Angel", "Early-Stage VC", and "Seed Round". We include dummies indicating the type of deal of a focal investment.

Finally, we include fixed effects for country locations of entrepreneurial firms, ϕ_c, fixed effects for the year of founding, τ_e, and fixed effects for year of the investment, τ_t. Standard errors are clustered at the country location of start-up companies.

We present the estimation results in Table 3. To examine the contingency described in H1, we consider a sample consisting of ventures founded by research-oriented founders with expertise in the same domain as the ventures' industry sector as well as their matched ventures without research-oriented founders onboard. Research-oriented founders experience significantly better outcomes than non-research-oriented founders (Columns 3 and 4). Based on the outcome of success (good exit), research-oriented founders experience an absolute increase of 3.9% (1.8%) in the likelihood of achieving an exit compared to non-research-oriented founders, holding all other variables constant. Such effects are economically significant given the mean rates for success being 26% and for good exit being 8% in the sample.

As a benchmark, we also report the estimation results using two different samples: the entire sample of matched pairs of ventures with and without research-oriented founders (Columns 1 and 2), and a sample of ventures with research-oriented founders having expertise in domains different from the ventures' industry sector and their matched ventures without research-oriented founders (Columns 5 and 6). Research-oriented founders experience significantly better venture performance compared to their matched non-research-oriented founders only when we use *Success* as the measure for venture outcome (Column 1). Though research-oriented founders continue to show higher rates of experiencing a good exit event than non-research-oriented founders, the effect is not statistically significant (Column 2), a result which might be compared to those reported in Roche, et al. (2020). The key finding in that article is no significant performance difference between research- versus non-research-oriented founders, though the authors do not adopt a contingency-based analysis (nor an empirical matching approach), as we do here. Furthermore, when research-oriented founders start a venture in domains different from their expertise', they perform *worse* than non-research-oriented founders, holding all other variables constant.

We examine the robustness of the results using a sample conditioned on identifying research-oriented founders as those with greater than one academic publication as recorded in

MAG. Among the founders for whom we can link publication records from MAG, about 25% have merely one publication record by the time of the first investment received by their venture, suggesting a skewed distribution of publications. As a robustness test, we re-run the analysis by adopting a more stringent definition for research-oriented founders (i.e., with more than one publication). The results remain intact (see Table A.4 in the Appendix for details).

4.2 Effects of research productivity and quality on venture outcomes

In this section, we restrict our attention to venture outcomes experienced by research-oriented founders and test H2. We now consider 3,836 ventures all founded by research-oriented entrepreneurs that raised their first equity investments between 2007 and 2015. We modify equation (1) and estimate the following equation:

$$Y_{evct} = \beta_0 + \beta_1 Pub_{et} + \beta_2 C_{et} + \beta_3 X_{vt} + \phi_c + \tau_t + \tau_e + \epsilon_{evct} \qquad (2)$$

where e, v, c, and t index the start-up company, lead investor, country, and investment year, respectively. Like our analysis in testing H1, we use two alternative measures for success as the dependent variable (Y_{evct}): *Success* and *Good Exit.* Our variable of interest, Pub_{et}, spans measures for the quality or quantity of academic publications authored by research-oriented founders. Specifically, we consider the following measures at the time of the first investment: (i) the logged number of total citations received by all the papers published by members of the founding team; (ii) highest H-index (defined as the maximum value, *h*, such that the given author has published at least *h* articles, each of which have been cited at least *h* times) among all founding team members; and (iii) logged value of the total number of papers published by all members on the founding team. We include the same set of controls as explained in Section 4.1 and cluster standard errors at the country level of start-up companies.

Table 4 reports the estimation results. Overall, we find results supporting H2, that the quality of prior research work by founders is positively related to the likelihood of experiencing a successful exit event. Specifically, the two measures that reflect quality of research work, H-index and forward citations, are positively and significantly associated with the measures of successful exits. However, research quantity, as assessed by the number of published papers, is not significantly related to the occurrence of either type of successful exit event. The results suggest that quality rather than quantity of research matters more to venture outcomes of research-oriented founders. Furthermore, these results are likely driven by research-oriented founders starting ventures in their domains of research expertise. In Table A.5 in the Appendix, we report results from analyzing two groups of ventures separately: those in the same domains with expertise of their research-oriented founders (Panel A), and those in a domain different from expertise of their research-oriented founders (Panel B). Research quality is positively and significantly related to likelihoods of exits only for ventures that are in the same domains with expertise of their research-oriented founders.

4.3 Effects of extended "team" complementary human capital

We now turn to examining complementarity of investor expertise as well as human capital of members other than the research-oriented founders on the team. In testing H3, we further restrict our sample to include ventures founded by STEM-background researchers and the matched ventures that were not founded by research-oriented individuals. First, based on educational major/concentration information, we identify if any members on the founding team have educational backgrounds in business, economics, or law. We construct a binary variable, *business/law edu*, that equals to one if at least one member on the founding team holds degrees in the areas of business, economics, or law.

Second, we evaluate whether a lead investor on the focal deal is a specialist-investor. As suggested in the literature, investors usually focus on investment opportunities in selected industries and develop expertise in screening and adding value to start-up companies in their specialized fields.[23] Therefore, we assess if founders can tap into the expertise of those specialist-investors when starting ventures in domains different from their own research fields for venture success. In identifying a specialist, we start from calculating the share of total number of deals invested in a focal industry group over the total number of deals previously made across all industry groups by an investor. We then identify an investor as a *Specialist Investor* if their previous investments in a focal industry group exhibits a level of at least 47% of their overall investments, which corresponds to the 75th percentile of industry concentration in the data.

In testing H3a and H3b, we investigate whether human capital of the extended team (i.e., founding team members and investors) complements the human capital of research-oriented founders. We estimate the following equation:

$$Y_{evct} = \gamma_0 + \gamma_1 ResearchFounder_e + \gamma_2 ResearchFounder_e \times H + \gamma_3 H + \gamma_4 C_{et} + \gamma_5 X_{vt} + \phi_c + \tau_t + \tau_{\ e} + \epsilon_{evct} \qquad (3)$$

[23] For example, Gompers et al. (2020) surveyed 885 individual venture capitalists and about 61% responded that they focus on investing in no more than two industries.

H includes the two variables that measure human capital of the "extended team": *business/law edu* and *Specialist Investor*. The interaction term captures whether those two variables further modulate the effect of having a research-oriented founder.

We report the results from estimating equation 3 in Table 5. Results in columns (5) and (6) are based on a sample of ventures started by research-oriented founders in industry sectors different from their domains of expertise and the matched ventures of non-research-oriented ventures. We also report estimation from using a sample of ventures started by research-oriented founders in industry sectors within their domains of expertise and the matched ventures of non-research-oriented ventures (Columns 3 and 4). We find supportive evidence for H3a, that having a founding team with complementary skills (e.g., business, marketing, law) will help compensate for venture outcomes, as suggested by the positive and significant estimated coefficient on the interaction term. In fact, if assembling a founding team with educational backgrounds in the areas of business, economics or law, research-oriented founders will experience significantly higher likelihoods of venture exits compared to non-research-oriented founders if they start ventures in domains different from their expertise. However, this complementarity does not hold when research-oriented founders launch a venture in the same domain of their expertise (Columns 3 and 4).

On the other hand, we do not find supportive evidence for H3b. The coefficient on the interaction between *ResearchFounder* and *Specialist-Investor* is estimated to be negative but not statistically different than zero (Columns 5 and 6). An ex-post analysis reveals that a complementarity between *ResearchFounder* and having a *Specialist-Investor* as lead investor arises when research-oriented founders start ventures in the same domains of their expertise. This is indicated by the positive coefficient on the interaction shown in Columns 3 and 4. This suggests a potential reinforcing effect of the knowledge and resources of founders and investors within the same domain, and is an intriguing phenomenon for future work.

5. Conclusion and Future Directions

We add to the small number of studies examining STEM "research-oriented founders" by conceptualizing the circumstances under which such a background is associated with venture outperformance (as compared to otherwise observationally similar enterprises). We spotlight an important contingency which has been neglected in the prior literature: whether the focal enterprise development is within or outside the research-oriented founders' domain of expertise, as each circumstance is related to a different configuration of benefits associated with complementary co-founders and investors. Our empirical sample is much broader relative to prior efforts, and our data matching approach may help mitigate the confounding role of unobserved differences. These improvements, together with our contingency-based theorizing, may help reconcile the results reported by Roche et al. (2020) with the broader human capital and entrepreneurship literature.

This is significant given that both educational and managerial policy depend on the relationship between research-oriented founders' skills and venture performance. On the former, because of the increasingly large mismatch between the number of STEM graduates produced and the market need for full-time academic positions, understanding the re-deployability of knowledge and skills for alternative careers including venture development is important.[24] In the latter domain, holders of financial and human resources (including co-founders and venture joiners) benefit from understanding the success patterns of research-oriented founders.

We would also like to discuss two limitations of the study (which present opportunities for future research). First, we caution that the relations we document should not be interpreted as causal because our empirical design does not involve a randomized or natural experiment. We speculate that a randomization design would be hard to achieve given the large investments graduate programs typically make in training and placing their students. Consequently, while we aim to mitigate the confound of unobserved selection through an empirical matching approach, unobserved and/or unmeasured processes prevent a causal interpretation. A second limitation is that our understanding of the "why" answers to the empirical patterns we document are currently limited. Future work would ideally delve deeper into the potential mechanisms driving the results, for example by better understanding whether research oriented founders are better able to identify entrepreneurial opportunities or are alternatively better at attracting resources to join their venture development efforts.[25] Our hope is that these and other follow-on questions will be the subject of future study in this emerging, but important, domain of research.

---

[24] While Camuffo et al (2020) suggest that an educational intervention highlighting a scientific approach to venture development is beneficial, it is unknown how such training compares to the likely scientific method-ingrained approach undertaken by STEM-trained venture founders. A separate issue of potential importance to educational institutions is the possible (geographically local) spillovers that research-oriented founders may have (a well-known example is the case of Google, which has a market value of $1.5T as of August 2022).

[25] As preliminary evidence of potential mechanisms, we find that within-domain research-oriented founders tend to obtain investment from more experienced and network-central VCs, with the economic relation lower for outside-domain research-oriented founders (Appendix A.6). We suspect other processes are salient, however.

References

Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2020). Age and high-growth entrepreneurship. *American Economic Review: Insights*, *2*(1), 65-82.

Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*, *111*(16), 5773-5777.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9-49.

Blank, R., Ronald J. Daniels, R. J., Gilliland, G., Gutmann, A., Hawgood, S., Hrabowski, F. A., Pollack, M. E., Price, V., Reif, L. R., and Schlissel, M. S (2017). "A new data effort to inform career choices in biomedicine." *Science* 358(6369): 1388-1389.

Brüderl, J., Preisendörfer, P., & Ziegler, R. (1992). Survival chances of newly founded business organizations. *American Sociological Review*, 227-242.

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, *66*(2), 564-586.

Cassar, G. (2014). Industry and startup experience on entrepreneur forecast performance in new firms. *Journal of Business Venturing*, *29*(1), 137-151.

Christensen, C. M. (1993). The rigid disk drive industry: A history of commercial and technological turbulence. *Business History Review*, *67*(4), 531-588.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95-S120.

Colombo, M. G., & Grilli, L. (2005). Founders' human capital and the growth of new technology-based firms: A competence-based view. *Research Policy*, *34*(6), 795-816.

Colombo, M. G., & Grilli, L. (2010). On growth drivers of high-tech start-ups: Exploring the role of founders' human capital and venture capital. *Journal of Business Venturing*, *25*(6), 610-626.

Colombo, M. G., & 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.

Druilhe, C., & Garnsey, E. (2004). Do academic spin-outs differ and does it matter?. *The Journal of Technology Transfer*, *29*(3), 269-285.

Ewens, M., Gorbenko, A., & Korteweg, A. (2022). Venture capital contracts. *Journal of Financial Economics*, *143*(1), 131-158.

Ewens, M., & Marx, M. (2018). Founder replacement and startup performance. *The Review of Financial Studies*, *31*(4), 1532-1565.

Ewens, M., Rhodes-Kropf, M., & Strebulaev, I. A. (2016). Insider financing and venture capital returns. Working paper

Gompers, P. A., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions?. *Journal of Financial Economics*, *135*(1), 169-190.

Helfat, C. E., & Lieberman, M. B. (2002). The birth of capabilities: market entry and the importance of pre-history. *Industrial and Corporate Change*, *11*(4), 725-760.

Hochberg, Y. V., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, 62(1), 251-301.

Hsu, D. H. (2004). What do entrepreneurs pay for venture capital affiliation?. *The Journal of Finance*, *59*(4), 1805-1844.

Hsu, D. H. (2007). Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy*, *36*(5), 722-741.

Karlan, D., & Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics*, *93*(2), 510-527.

Klepper, S. (2001). Employee startups in high-tech industries. *Industrial and Corporate Change*, *10*(3), 639-674.

Larson, R. C., Ghaffarzadegan, N., & Xue, Y. (2014). Too many PhD graduates or too few academic job openings: The basic reproductive number R0 in academia. *Systems Research and Behavioral Science*, *31*(6), 745-750.

Langin, K. (2019). In a first, US private sector employs nearly as many Ph. Ds as schools do. *Science*, *12*.

Levin, S. G., & Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 114-132.

Levin, M., Krawczyk, S., Bethard, S., & Jurafsky, D. (2012). Citation-based bootstrapping for large-scale author disambiguation. *Journal of the American Society for Information Science and Technology*, *63*(5), 1030-1047.

Martin, B. C., McNally, J. J., & Kay, M. J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, *28*(2), 211-224.

Marx, M., & Hsu, D. H. (2022). Revisiting the entrepreneurial commercialization of academic science: Evidence from "Twin" discoveries. *Management Science*, *68*(2), 1330-1352.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research Policy*, *31*(8-9), 1389-1403.

Oosterbeek, H., Van Praag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, *54*(3), 442-454.

Roach, M., & Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. *Research Policy*, *39*(3), 422-434.

Roach, M., & Sauermann, H. (2015). Founder or joiner? The role of preferences and context in shaping different entrepreneurial interests. *Management Science*, *61*(9), 2160-2184.

Roberts, E. B. (1991). *Entrepreneurs in high technology: Lessons from MIT and beyond*. Oxford University Press.

Roche, M. P., Conti, A., & Rothaermel, F. T. (2020). Different founders, different venture outcomes: A comparative analysis of academic and non-academic startups. *Research Policy*, *49*(10), 104062.

Rothaermel, F. T., Agung, S. D., & Jiang, L. (2007). University entrepreneurship: a taxonomy of the literature. *Industrial and Corporate Change*, *16*(4), 691-791.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. *American Sociological Review*, 195-222.

Shah, S. K., & Pahnke, E. C. (2014). Parting the ivory curtain: understanding how universities support a diverse set of startups. *The Journal of Technology Transfer*, *39*(5), 780-792.

Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, *11*(4), 448-469.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312), aaf5239.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B. J., & Wang, K. (2015, May). An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web* (pp. 243-246).

Tripsas, M. (1997). Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry. *Strategic Management Journal*, *18*(S1), 119-142.

Unger, J. M., Rauch, A., Frese, M., & Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing*, *26*(3), 341-358.

Vohora, A., Wright, M., & Lockett, A. (2004). Critical junctures in the development of university high-tech spinout companies. *Research Policy*, *33*(1), 147-175.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., Rogahn, R. (2019). A review of microsoft academic services for science of science studies. *Frontiers in Big Data*, *2*, 45.

Wang, K., Shen, Z., Huang, C., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396-413.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039.

Table 1. Start-up companies with research-oriented founders

The sample contains 3,836 start-up companies that were founded by research-oriented founders and received a first external equity investment from 2007 to 2015 worldwide. Research-oriented founders are identified based on the algorithm introduced in Section 3.2 and are linked to MAG publication records. Panel A tabulates the number of publication records in MAG for research-oriented founders in the STEM fields. Panel B counts the number and share of ventures with research-oriented founders by industry sectors and further differentiates whether research-oriented founders launch their ventures in the same domain as their fields of research expertise. In Appendix C, we provide a detailed description to the matching relationship between industry sectors in Pitchbook and level-0 fields of publications in MAG.

A. Number of publications by level-0 STEM fields in MAG

Level-0 Fields	No. of Publications	Percent
Biology	22,352	17.3
Chemistry	15,753	12.2
Computer science	29,436	22.8
Engineering	6,598	5.1
Environmental science	1,105	0.9
Geography	767	0.6
Geology	609	0.5
Materials science	14,514	11.3
Mathematics	4,511	3.5
Medicine	21,975	17.1
Physics	7,910	6.1
Psychology	3,347	2.6
Total	128,877	100.0

B. Industry sectors of ventures founded by research-oriented founders

	Entire sample		Same as domain of research-founders' expertise		Different from domain of research-founders' expertise	
	(1)	(2)	(3)	(4)	(5)	(6)
	Freq.	Percent	Freq.	Percent	Freq.	Percent
Business Products and Services (B2B)	379	9.88	N/A	N/A	379	22.40
Consumer Products and Services (B2C)	434	11.31	N/A	N/A	434	25.65
Energy	116	3.02	61	2.82	55	3.25
Financial Services	48	1.25	N/A	N/A	48	2.84
Healthcare	1259	32.82	1180	55.04	79	4.67
Information Technology	1530	39.89	843	39.32	687	40.60
Materials and Resources	70	1.82	60	2.80	10	0.59
Total	3836	100	2144	100	1692	100

Table 2. Summary statistics

A. Sample of matched research-oriented ventures and non-researcher founders

We perform exact matching between ventures founded by research-oriented founders and by non-research-oriented founders. We match on the following criteria: country location, industry group (41 groups), year of investment, and year of company foundation.

	Mean	Median	Min	Max	S.D.
Success	0.26	0.00	0.00	1.00	0.44
Good exit	0.08	0.00	0.00	1.00	0.28
Deal size($mil)	3.26	0.86	0.00	350.00	9.87
Team size	2.36	2.00	1.00	12.00	1.16
Female founder	0.18	0.00	0.00	1.00	0.39
Serial founder	0.13	0.00	0.00	1.00	0.34
Leader round exp	80.34	23.00	0.00	1356.00	157.41
No of investor	2.71	2.00	1.00	44.00	3.12
Observations	5540				

B. Sample of research-oriented founders' ventures

The sample contains 3,836 start-up companies founded by research-oriented founders and receiving first external equity investment from 2007 to 2015 worldwide. Research-oriented founders are identified based on the algorithm introduced in Section 3.2 and are linked to MAG publication records.

	Mean	Median	Min	Max	S.D.
Citation	1486.42	69.50	0.00	2.2e+05	7177.60
Paper sum	43.65	7.00	0.00	2186.00	124.66
H-index	7.99	3.00	0.00	202.00	14.63
Success	0.28	0.00	0.00	1.00	0.45
Good exit	0.10	0.00	0.00	1.00	0.30
Deal size	3.51	1.01	0.00	218.27	8.90
Team size	2.60	2.00	1.00	18.00	1.28
Female founder	0.19	0.00	0.00	1.00	0.39
Serial founder	0.14	0.00	0.00	1.00	0.35
Leader round exp	80.38	24.00	0.00	1392.00	151.03
No of investor	2.67	2.00	1.00	39.00	2.84
Observations	3836				

Table 3. Venture outcomes for research-oriented versus non-research-oriented founders

This table presents results from estimating equation (1) using a linear probability model (LPM) using three sets of observations: the entire sample of all matched ventures (Columns 1 and 2), ventures founded by research-oriented founders in the same domain as their expertise and the matched ventures founded by non-research-oriented founders (Columns 3 and 4), and a sample of ventures founded by research-oriented founders in a domain different from their expertise and the matched ventures founded by non-research-oriented founders (Columns 5 and 6). Detailed explanations of controls can be found in Section 4.1. Standard errors are clustered by country locations.

	(1)	(2)	(3)	(4)	(5)	(6)
	All observations		Research-founders in the same domain		Research-founders in a different domain	
	Success	Good exit	Success	Good exit	Success	Good exit
Research-founder	0.0184**	0.00359	0.0398***	0.0183***	-0.00798	-0.0157***
	(0.00843)	(0.00337)	(0.0102)	(0.00398)	(0.00610)	(0.00402)
Ln(deal size)	0.0733***	0.0693***	0.0820***	0.0868***	0.0486***	0.0249***
	(0.00958)	(0.00411)	(0.00763)	(0.00382)	(0.0158)	(0.00873)
Ln(team size)	0.0523***	0.0400***	0.0568***	0.0453***	0.0493***	0.0382***
	(0.0116)	(0.00711)	(0.0180)	(0.0112)	(0.0130)	(0.00829)
Female founder	-0.0130	-0.0120***	-0.0102	-0.00730	-0.0111	-0.0136***
	(0.00960)	(0.00361)	(0.00907)	(0.00432)	(0.0127)	(0.00457)
Serial founder	-0.00111	0.0265**	-0.0273	0.0283	0.0296*	0.0187**
	(0.0115)	(0.01000)	(0.0294)	(0.0171)	(0.0158)	(0.00775)
Ln(leader exp)	0.0186***	0.00558**	0.0211***	0.00805***	0.0154***	0.00256**
	(0.00414)	(0.00226)	(0.00504)	(0.00282)	(0.00313)	(0.00120)
Ln(no. of investors)	0.0450***	0.00359	0.0421***	0.00225	0.0525***	0.0107*
	(0.00370)	(0.00293)	(0.00375)	(0.00435)	(0.00632)	(0.00531)
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Adjusted R^2	5540	5540	3062	3062	2478	2478
Observations	0.122	0.109	0.129	0.132	0.107	0.063

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Table 4. Research-oriented founders: effects of research quality and quantity on venture outcomes

This table presents results for testing H2 using a sample consisting of all the ventures founded by research-oriented founders. We use a linear probability model (LPM) to estimate equation (2). Three alternative measures of research quality or quantity by the time of the first-round investments are considered: H-index (Columns 1 and 2), logarithm value of citations received (Columns 3 and 4), and logarithm of number of papers published (Columns 5 and 6). A detailed explanation of controls can be found in Section 4.2. Standard errors are clustered by country locations.

	(1) Success	(2) Good exit	(3) Success	(4) Good exit	(5) Success	(6) Good exit
H-index	0.000945***	0.000789***				
	(0.000298)	(0.0000900)				
Ln(no. of citations)			0.00410***	0.00311***		
			(0.00136)	(0.00111)		
Ln(no. of papers)					0.00341	0.00276*
					(0.00249)	(0.00150)
Ln (deal size)	0.0860***	0.0743***	0.0865***	0.0748***	0.0875***	0.0756***
	(0.00986)	(0.00594)	(0.00956)	(0.00615)	(0.00960)	(0.00609)
Ln (team size)	0.0712***	0.0320***	0.0732***	0.0339***	0.0740***	0.0344***
	(0.0202)	(0.0111)	(0.0184)	(0.0102)	(0.0192)	(0.0107)
Female founder	-0.00716	-0.0151***	-0.00794	-0.0158***	-0.00803	-0.0159***
	(0.0130)	(0.00544)	(0.0131)	(0.00546)	(0.0131)	(0.00547)
Serial founder	-0.0101	0.0346***	-0.00857	0.0361***	-0.00754	0.0368***
	(0.00877)	(0.0110)	(0.00918)	(0.0115)	(0.00913)	(0.0115)
Ln (lead exp)	0.0165**	0.00248	0.0166**	0.00255	0.0166**	0.00258
	(0.00703)	(0.00509)	(0.00699)	(0.00506)	(0.00706)	(0.00511)
Ln (no. of investors)	0.0372***	0.0135**	0.0365***	0.0128**	0.0365***	0.0128**
	(0.00844)	(0.00618)	(0.00845)	(0.00630)	(0.00845)	(0.00629)
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	3836	3836	3836	3836	3836	3836
Adjusted R^2	0.117	0.121	0.117	0.120	0.116	0.119

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 5. Effects of extended "team" complementary human capital

This table presents results from estimating equation (3) using a linear probability model (LPM). There are three sets of observations: the entire sample of all matched ventures (Columns 1 and 2), ventures founded by research-oriented founders in the same domain as their expertise and the matched ventures founded by non-research-oriented founders (Columns 3 and 4), and a sample of ventures founded by research-oriented founders in a domain different from their expertise and the matched ventures founded by non-research-oriented founders (Columns 5 and 6). Detailed explanation of controls can be found in Section 4.1. Standard errors are clustered by country locations.

	(1)	(2)	(3)	(4)	(5)	(6)
	All observations		Research-founders in the same domain		Research-founders in a different domain	
	Success	Good exit	Success	Good exit	Success	Good exit
Research-founder	-0.000637	-0.00150	0.0259**	0.0114**	-0.0358**	-0.0220***
	(0.00750)	(0.00396)	(0.0110)	(0.00501)	(0.0160)	(0.00695)
Research-founder*business/law edu	0.0301***	-0.00200	0.0169	-0.00653	0.0569**	0.0153*
	(0.00778)	(0.00555)	(0.0111)	(0.00858)	(0.0258)	(0.00867)
Business/law edu	-0.0263***	-0.00590	-0.0244*	-0.0206***	-0.0313	0.00934
	(0.00836)	(0.00415)	(0.0133)	(0.00356)	(0.0244)	(0.00641)
Research-founder*Specialist-investor	0.0179	0.0214***	0.0113	0.0183***	-0.000426	0.00206
	(0.0116)	(0.00400)	(0.0103)	(0.00623)	(0.0256)	(0.0107)
Specialist-investor	0.0327**	0.0241***	0.0460***	0.0206***	0.00810	0.0213
	(0.0137)	(0.00717)	(0.00579)	(0.00602)	(0.0361)	(0.0133)
Ln(deal size)	0.0725***	0.0683***	0.0801***	0.0856***	0.0487***	0.0252***
	(0.00991)	(0.00410)	(0.00774)	(0.00394)	(0.0153)	(0.00879)
Ln(team size)	0.0568***	0.0434***	0.0623***	0.0546***	0.0502***	0.0324***
	(0.0106)	(0.00785)	(0.0190)	(0.0117)	(0.00925)	(0.00801)
Female founder	-0.0126	-0.0114***	-0.00845	-0.00614	-0.0124	-0.0130**
	(0.00918)	(0.00356)	(0.00861)	(0.00414)	(0.0135)	(0.00479)
Serial founder	-0.00224	0.0260**	-0.0290	0.0287*	0.0289*	0.0181**
	(0.0108)	(0.00964)	(0.0290)	(0.0164)	(0.0159)	(0.00791)
Ln(lead exp)	0.0176***	0.00481**	0.0199***	0.00729**	0.0153***	0.00238**
	(0.00370)	(0.00211)	(0.00496)	(0.00273)	(0.00249)	(0.00112)

Ln(no. of investors)	0.0442***	0.00313	0.0412***	0.00228	0.0520***	0.00979*
	(0.00351)	(0.00278)	(0.00361)	(0.00425)	(0.00550)	(0.00500)
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	5540	5540	3062	3062	2478	2478
Adjusted R^2	0.123	0.111	0.130	0.134	0.106	0.060

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Appendix

Appendix A. Additional Tables

Table A.1. Articles in MAG

Panel A. Percentage of MAG articles with different levels of fields available

MAG Field Level	Percentage of articles with the field assignment available
0	98.78%
1	94.54%
2	77.90%
3	61.66%
4	26.07%
5	6.85%

Panel B. Tabulation of articles (patents excluded) by level-0 fields in MAG

Level-0 Fields	Counts of articles
Art	8,649,106
Geology	6,332,204
Geography	5,529,948
Economics	2,905,504
Computer science	24,442,356
Philosophy	4,155,656
Biology	14,203,892
Political science	8,169,567
Sociology	4,560,973
Engineering	14,553,116
Chemistry	19,185,777
Environmental science	5,916,113
Medicine	29,693,616
Business	5,222,129
History	3,370,175
Psychology	8,429,090
Mathematics	6,709,081
Physics	10,380,853
Materials science	28,646,481

Table A.2. Country locations of ventures with research-oriented founders

Headquarter Location	Freq.	Percent	Cum.
United States	2,525	65.82	65.82
United Kingdom	272	7.09	72.91
Canada	142	3.70	76.62
Israel	93	2.42	79.04
India	81	2.11	81.15
China	74	1.93	83.08
France	70	1.82	84.91
Switzerland	64	1.67	86.57
Germany	55	1.43	88.01
Ireland	53	1.38	89.39
Other	407	10.61	100.00
Total	3,836		

Table A.3. Average marginal effects from estimating Equation 1 by Probit (replication of Table 3 using Probit estimation)

This table presents results from estimating equation (1) by Probit using three sets of observations: the entire sample of all matched ventures (Columns 1 and 2), ventures founded by research-oriented founders in the same domain as their expertise and the matched ventures founded by non-research-oriented founders (Columns 3 and 4), and a sample of ventures founded by research-oriented founders in a domain different from their expertise and the matched ventures founded by non-research-oriented founders (Columns 5 and 6). Detailed explanations of controls can be found in Section 4.1. Average marginal effects are reported and standard errors are clustered by country locations.

	(1)	(2)	(3)	(4)	(5)	(6)
	All observations		Research-founders in the same domain		Research-founders in a different domain	
	Success	Good exit	Success	Good exit	Success	Good exit
Research-founder	0.0195**	0.00345	0.0390***	0.0190***	-0.00436	-0.0176***
	(0.00776)	(0.00323)	(0.00894)	(0.00401)	(0.00636)	(0.00408)
Ln(deal size)	0.0620***	0.0430***	0.0705***	0.0575***	0.0438***	0.0202***
	(0.0100)	(0.00372)	(0.00900)	(0.00371)	(0.0143)	(0.00612)
Ln(team size)	0.0520***	0.0383***	0.0532***	0.0419***	0.0495***	0.0421***
	(0.0131)	(0.00797)	(0.0176)	(0.0118)	(0.0146)	(0.00817)
Female founder	-0.0148	-0.0207***	-0.0161	-0.0203***	-0.00863	-0.0192***
	(0.0110)	(0.00512)	(0.00989)	(0.00619)	(0.0156)	(0.00719)
Serial founder	0.000852	0.0224***	-0.0267	0.0204	0.0311*	0.0198***
	(0.0107)	(0.00792)	(0.0301)	(0.0150)	(0.0163)	(0.00663)
Ln(lead VC exp)	0.0189***	0.00677***	0.0210***	0.00943***	0.0159***	0.00352**
	(0.00390)	(0.00219)	(0.00499)	(0.00285)	(0.00300)	(0.00139)
Ln(no. of investors)	0.0363***	0.00534**	0.0336***	0.00453	0.0415***	0.0120***
	(0.00311)	(0.00237)	(0.00373)	(0.00381)	(0.00541)	(0.00323)
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	5540	5540	3062	3062	2478	2478
Pseudo R^2	0.13	0.18	0.14	0.19	0.13	0.15

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Table A.4. Ventures with research-oriented founders publishing more than one paper

This table presents results from estimating equation (1) by linear probability model using a sample of research-founders with **more than one** published paper and their matched ventures. Detailed explanations of controls can be found in Section 4.1. Standard errors are clustered by country locations.

	(1)	(2)	(3)	(4)	(5)	(6)
	All observations		Research-founders in the same domain		Research-founders in a different domain	
	Success	Good exit	Success	Good exit	Success	Good exit
Research-founder	0.0244**	0.00406	0.0388***	0.0161***	0.00312	-0.0152***
	(0.0109)	(0.00420)	(0.0123)	(0.00564)	(0.00932)	(0.00389)
Ln(deal size)	0.0711***	0.0754***	0.0785***	0.0912***	0.0479**	0.0272***
	(0.0103)	(0.00417)	(0.00695)	(0.00522)	(0.0198)	(0.00730)
Ln(team size)	0.0636***	0.0495***	0.0660***	0.0580***	0.0639***	0.0414***
	(0.0117)	(0.00734)	(0.0185)	(0.0121)	(0.0162)	(0.0116)
Female founder	-0.00799	-0.00343	-0.00979	0.00473	-0.00174	-0.0121
	(0.00766)	(0.00391)	(0.0106)	(0.00535)	(0.00958)	(0.00735)
Serial founder	-0.0148	0.0134*	-0.0499*	0.00696	0.0315**	0.0130
	(0.0152)	(0.00709)	(0.0273)	(0.0165)	(0.0127)	(0.0146)
Ln(lead VC exp)	0.0153***	0.00367	0.0194***	0.00699*	0.00918***	-0.000865
	(0.00462)	(0.00282)	(0.00558)	(0.00343)	(0.00326)	(0.00178)
Ln(no. of investors)	0.0452***	0.00811**	0.0486***	0.0113**	0.0426***	0.0107
	(0.00370)	(0.00319)	(0.00518)	(0.00446)	(0.00514)	(0.00773)
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	4156	4156	2456	2456	1700	1700
Adjusted R^2	0.124	0.110	0.128	0.129	0.107	0.060

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Table A.5. Research-oriented founders: effects of research quality and quantity on venture outcomes by domains of expertise

This table presents results from estimating equation (2) using a linear probability model. Results are estimated from using two sets of observations: ventures founded by research-oriented founders in the same domains as their expertise (Panel A), and ventures founded by research-oriented founders in a domain different from their expertise (Panel B). Three alternative measures of research quality or quantity by the time of the first-round investments are considered: H-index (Columns 1 and 2), logarithm value of citations received (Columns 3 and 4), and logarithm of number of papers published (Columns 5 and 6). Detailed explanation of controls can be found in Section 4.2. Standard errors are clustered by country locations.

	(1) Success	(2) Good exit	(3) Success	(4) Good exit	(5) Success	(6) Good exit
Panel A. Ventures in the same domains with expertise of research-oriented founders						
H-index	0.00162***	0.00107***				
	(0.000317)	(0.0000879)				
Ln (no. of citations)			0.00692***	0.00432*		
			(0.00223)	(0.00238)		
Ln(no. of papers)					0.00844***	0.00598**
					(0.00293)	(0.00269)
Controls	Y	Y	Y	Y	Y	Y
Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	2118	2118	2118	2118	2118	2118
Adjusted R^2	0.138	0.136	0.136	0.135	0.135	0.135
Panel B. Ventures in a different domain from expertise of research-oriented founders						
H-index	-0.00142***	-0.000788***				
	(0.000387)	(0.000241)				
Ln (citation)			-0.000651	-0.000394		
			(0.00234)	(0.00134)		
Ln(paper sum)					-0.00483	-0.00351
					(0.00457)	(0.00309)
Controls	Y	Y	Y	Y	Y	Y

Deal Type FE	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y
Observations	1718	1718	1718	1718	1718	1718
Adjusted R^2	0.085	0.047	0.083	0.046	0.084	0.046

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Table A.6. Effects on receiving investments from a reputable VC firm

Note: This table reports results from OLS estimation. The dependent variable is the reputation level of the lead investor of the first funding round received by a venture. Three alternative reputation measures are considered: count of round experience, eigenvector centrality, and closeness centrality. The sample considers matched pairs of research-oriented and non-research-oriented ventures that received their first rounds between 2010 and 2015. The sample period is shorter than that used in our baseline analysis as we leave at least three years to observe VC experience and build VC reputation measures. The centrality measures are based on a three-year rolling window of syndication relationships among VC firms *worldwide*. Round exp counts the number of rounds a lead investor participated in since the year 2007 until the time of a focal round. Standard errors are clustered at country level of entrepreneurial firms.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All observations			Research-founders in the same domain			Research-founders in a different domain	
Reputation measure	Ln(round exp)	Eigenvector	Closeness	Ln(round exp)	Eigenvector	Closeness	Ln(round exp)	Eigenvector	Closeness
Research-founder	0.355***	0.0000850***	0.00000104***	0.391***	0.0000921***	0.00000114***	0.296***	0.0000756***	0.000000885***
	(0.0143)	(0.0000208)	(7.88e-08)	(0.0267)	(0.0000215)	(9.50e-08)	(0.0264)	(0.0000214)	(0.000000112)
Ln(team size)	0.182***	0.0000339***	0.000000835***	0.0761	0.0000223	0.000000709***	0.301***	0.0000389***	0.000000824***
	(0.0476)	(0.00000837)	(0.000000125)	(0.0609)	(0.0000148)	(0.000000186)	(0.0669)	(0.0000137)	(0.000000188)
Female founder	0.0497	-0.00000109	-3.18e-08	0.0354	0.00000706	-6.80e-08	0.0877*	-0.00000168	2.82e-08
	(0.0298)	(0.00000497)	(6.90e-08)	(0.0501)	(0.0000135)	(8.16e-08)	(0.0475)	(0.0000196)	(0.000000120)
Serial founder	0.318***	0.000137***	0.000000793***	0.197***	0.0000913***	0.000000368***	0.459***	0.000187***	0.00000132***
	(0.0240)	(0.0000148)	(7.95e-08)	(0.0436)	(0.0000122)	(0.000000131)	(0.0479)	(0.0000285)	(0.000000128)
Deal Type FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry Group FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Investment Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Founding Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4650	4650	4650	2530	2530	2530	2120	2120	2120
Adjusted R^2	0.130	0.122	0.335	0.138	0.145	0.380	0.125	0.098	0.286

Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$

Appendix B. Linking founders in Pitchbook to authors in MAG

Author name disambiguation in MAG

MAG uses machine learning techniques to identify and disambiguate authors of scholarly publications (Wang et al. 2020). In addition to processing publication records, MAG incorporates information crawled and extracted by Microsoft's search engine service (Bing) from the entire Web. The sources for such data mining include personal websites and public curricula vitae. Other indexing systems, such as Google Scholar, have been critiqued for potentially inflating author publication records (erroneously assigning more publication records to an author). By comparison, MAG follows a more conservative approach: "…publications bearing the same author name are not assigned to the same author node in MAG unless such assignments can exceed a 97% confidence threshold on the machine learning algorithm. (Wang et al. (2020))"

Built on what the MAG team has already done on author name disambiguation, we further conduct a comprehensive disambiguation following the approach by Sinatra et al. (2016). This procedure iteratively merges publications if they are likely to have been authored by the same individual, and two authors are considered to be the same individual if all of the following three conditions hold true.

1. The last names of the two authors are identical;
2. The initials of the first names and, when available, given names are the same. If the full first names and given names are present for both authors, they have to be identical; and
3. One of the following is true:
   - The two authors cite each other at least once.
   - The two authors share at least one co-author.
   - The two authors share at least one similar affiliation, measured by TF-IDF and cosine similarity, following the approach in Levin et al. (2012).

The process stops once there is no pair of authors to merge. After the disambiguation procedure, we identify 241,397,335 authors, corresponding to a 7.3% decrease in the number of authors.

We further evaluate the accuracy of our algorithm to disambiguate author names. Specifically, we sample two data sets: (i) the first one consists of 200 pairs of papers that were identified by MAG as belonging to two distinct authors, but were indicated by the above procedure as written by the same author; (ii) the second data set consists of 200 pairs of papers, each appears to share a same author name but was not identified to be written by the same author by either MAG or our algorithm. For each of these 400 pairs of papers in the two data sets, we manually verify whether they were authored by the same individual by searching the authors' homepages or Google Scholar profiles. Overall, we find the false positive rate (i.e., fraction of times the procedure suggests the pair of papers belonging to the same author, while they actually do not) to be 0.5% (1/200) and a false negative rate (i.e., fraction of times that the same individual is considered to be two distinct persons) of 0%.

Matching fields of publications in MAG to educational majors/concentrations in Pitchbook

The similarity calculation procedure is as follows.

First, we obtain semantic representation of an educational major/concentration or a publication field as an embedding vector using pre-trained GloVe word embeddings (Pennington et al. (2014)). GloVe is one of the state-of-the-art models for distributed word representation, learned from global word-word co-occurrences. The pre-trained embeddings we use are learned based on 840 billion tokens, with 300 dimensions.[26] Specifically, each word contained in a name of a major/concentration or a name of a publication field is represented by a GloVe embedding vector. In the event that more than one word comprises a name of fields or of majors/concentrations, we combine vectors of each word into one representation for a single name, by adopting the commonly-used max pooling technique (Shen et al. (2018)), which keeps the maximum values in each dimension of the vectors.

Next, we calculate cosine similarity for each possible combination of major-field pair and adopt a threshold value of 0.8 for identifying a match. As we aim to match a major/concentration to a level-0 publication field, if a major/concentration is found to be similar with a publication field of a level lower than 0, we match the major to the level-0 publication field that is linked to the lower-level field.

In the end, we find at least one level-0 field in MAG for 9,025 distinct entries of educational fields in Pitchbook. Note that for a major involving inter-disciplinary knowledge, more than one level-0 field may be matched.

[26] https://nlp.stanford.edu/data/glove.840B.300d.zip

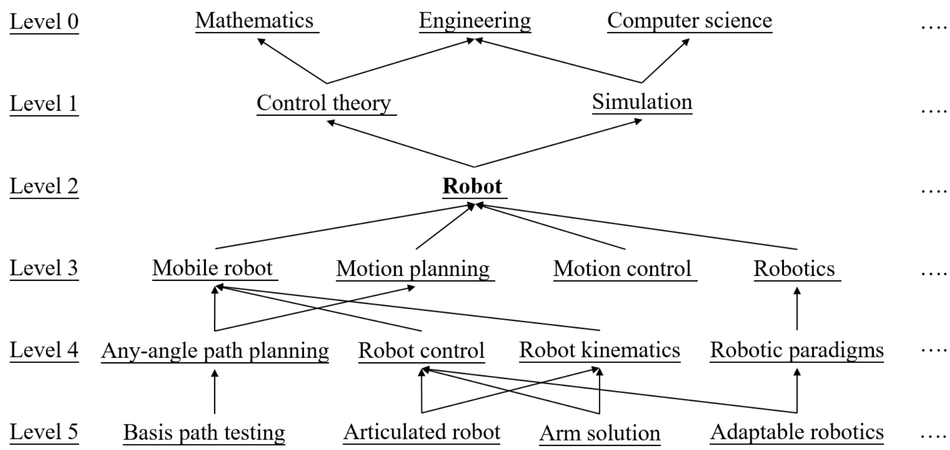

Figure B.1 MAG Taxonomy: An Example of All Publications in the Level-2 Field of "Robot"

Appendix C.

Matching relation between industry sectors and publication fields in MAG

Industry Sectors in PitchBook	Level-0 Fields in MAG
Information Technology	Computer Science
Healthcare	Biology, Medicine, Chemistry, Materials Science
Energy	Physics, Geology, Environmental Science, Chemistry, Materials Science.
Materials & Resources	Materials Science, Chemistry.