The Entrepreneurial Commercialization of Science: Evidence from "Twin" Discoveries

Matt Marx & David H. Hsu*

6 January 2019

Abstract: When are scientific advances translated into commercial products via startup formation? Although prior literature has offered several categories of answers, the commercial potential of a scientific advance is generally unobserved and potentially confounding. We assemble a sample of over 20,000 "twin" scientific discoveries in order to hold constant differences in the nature of the scientific advance, thereby allowing us to more precisely examine characteristics that predict startup commercialization. We find that teams of academic scientists whose former collaborators include "star" serial entrepreneurs are much more likely to commercialize their own discoveries via startups, as are more interdisciplinary teams of scientists.

Keywords: university technology transfer; entrepreneurship; technology commercialization; "twin" scientific discoveries.

* Marx: Boston University, mattmarx@bu.edu. Hsu: University of Pennsylvania, dhsu@wharton.upenn.edu. We thank Chris Ackerman, Rafael Castro, Andrea Contigiani, and Luming Yang for excellent research assistance and Christine Beckman for feedback. We also thank Guan-Cheng Li for data on patent-to-paper citations, Michael Ewens for his USPTO/VentureSource concordance, and Kyle Myers for his USPTO/Crunchbase/SBIR concordance. We acknowledge research support from Boston University and the Mack Center for Innovation Management at the University of Pennsylvania. Errors and omissions are ours.

1. Introduction

The technologies underlying some of the most successful companies in the world—including Google's PageRank search algorithm, E-Ink's "electronic paper", Mobileye's autonomous driving technology, RSA's cryptography algorithm, and Genentech's recombinant growth hormone—were discovered by academic scientists at universities who then commercialized them via startup formation. In 2017, the U.S. Association of University Technology Managers (AUTM) reported that over 1,000 startups were created from university intellectual property. Given increasing interest in academic technology commercialization (Sanberg, et al. 2014), we ask: what factors explain which scientific advances are translated into commercial products via startup formation?

New ventures as vehicles for commercializing academic technologies are important for at least three reasons. First, young firms are disproportionately involved in job growth (Haltiwanger, Jarmin & Miranda, 2013). Second, new ventures commercializing academic advances tend to co-locate near pioneering academic staff, thus contributing to regional economic development (Zucker, Darby & Brewer, 1998). Third, engaging the original scientists for on-going development may be critical to realizing the full potential of embryonic technologies (Jensen & Thursby, 2001).

Two prior literatures provide candidate answers to a general form of our research question. One stream highlights the role of entrepreneurial opportunity recognition, which in turn may be influenced by individual experience and the occupational or training environment in which the scientist is embedded. A second stream instead stresses financial and knowledge resource munificence in the institutional or local entrepreneurial ecosystem. One common factor clouding inference in the typical study across these two streams of work, however, is the difficulty of controlling for differences at the technological discovery level. Academic discoveries, in particular, can have varied "latent" commercial potential which can be difficult to discern (unknown even to the participants, much less to us as researchers).

Our empirical design builds on the Bikard & Marx (2018) method of assembling "twin" scientific discoveries. We scale up this effort considerably and study the likelihood that only one of two articles reporting the same scientific discovery is commercialized via a startup. Doing so helps to hold constant technological differences, enabling us to examine the empirical salience of two novel team-based variables within the entrepreneurial opportunity-based literature in predicting startup commercialization: 1) prior "star" commercialization peer effects and 2) more interdisciplinary collaborators.

2. Related literature

As noted, a key challenge that both the entrepreneurial opportunity recognition and entrepreneurial ecosystem literatures grapple with is the suitability of different scientific advances and technologies to be translated into commercial applications.[1] For example, Azoulay, Ding & Stuart (2007) address the related research question of "who patents?" among academic scientists. They conclude that patenting is a function of scientific opportunities (patenting is often observed after a flurry of scientific publications); furthermore, these same authors construct a measure of "latent commercializability," which predicts faculty patenting. The measure is based on keywords in scientists' publications which overlap with words which had previously been used in patent applications. Closer to the setting of commercialization via new venture formation, Shane (2001) finds differences in patent characteristics (such as measured "importance" and "scope") correlate with university technologies which are the subject of startup formation.[2]

A more fundamental problem is that the literature on academic commercialization frequently characterizes academy-originated technologies as "embryonic" (Jensen & Thursby, 2001), which compounds the difficulty of ascertaining and measuring latent commercializability. The inference issue therefore becomes more than one simply of measurement. Instead, if participants and decision-makers themselves are uncertain of the latent commercializability of a given technology, the problem of unobservability looms much larger. Even holding a given technology constant, there can be dramatic differences in entrepreneurial opportunity recognition (and hence commercialization attempts). As an example, Shane (2000) analyzes the case of eight sets of entrepreneurs recognizing different opportunities in response to a single invention, Three-D Printing™. His results illustrate that entrepreneurship reflects differences in information about opportunities, and that individual differences influence commercialization paths. In our context, differences among scientists who come up with "twin" discoveries may help to explain variation in entrepreneurial outcomes.

Not only can individual differences impact entrepreneurial opportunity recognition, several studies also stress the important role of peer influence. Nanda & Sorenson (2010) find a positive relationship between exposure to peers in the workplace with entrepreneurial experience and the likelihood of venture

[1] The entrepreneurial ecosystem literature explaining commercialization analyzes the role of the resource environment, particularly venture capital. The outcome of interest is typically centered on economic activity (venture starts, employment growth, etc.) at the regional level, with measures of venture capital activity as a regressor (e.g., Stuart & Sorenson, 2003; Samila & Sorenson, 2011). Latent commercializability is also an issue in this literature; however, due to the differing levels of analysis (regional economic outcomes versus individual scientific discoveries), it is difficult to directly compare results from this literature with ours.

[2] Note that there is a selection process associated with patenting, particularly in the university context, as the decision to pursue a patent reflects a prior assessment about commercialization prospects. In addition, it is not possible to determine the overall goodness-of-fit of patent characteristics in the reported Cox models of the hazard of new venture formation.

founding. In an empirical setting closer to ours, Stuart & Ding (2006) similarly find that academic life scientists who engage in commercial science (via firm founding or joining a scientific advisory board) have more coauthors with a commercial orientation. These studies all interpret peer effects as a social means of information transmission, which in turn impact the likelihood of startup-based commercialization.

We build on this peer-effects channel, shifting the unit of the analysis from an individual worker to the scientific discovery. This change is essential given our research question but more fundamentally because most discoveries are produced by *teams* of scientists (Wuchty, Jones, & Uzzi, 2007). Specifically, we propose that *scientists' prior affiliation with "star commercializers" – those in the extreme right tail of commercial activity, will likely have a positive impact on entrepreneurial commercialization attempts, holding the scientific discovery constant*. This is because such exposure is likely to be highly salient – as in the case of MIT professor Robert Langer, who has helped spawn over 40 startups and has been awarded over 1,000 patents (Krass, 2018).

We also investigate the influence of scientific-team composition on commercialization outcomes. A growing literature on team composition and performance suggests that more interdisciplinary teams can outperform (under certain circumstances) those of higher ability (Hong & Page, 2004). Leahy, Beckman, & Stanko (2017) find that even though interdisciplinary research is more difficult to publish, it garners more attention. *In the context of entrepreneurial commercialization, we expect more interdisciplinary teams are more likely to collectively recognize commercial applications, holding constant the technical advance.*

3. Empirical approach

As discussed above, a primary challenge in assessing when academic researchers commercialize their discoveries via startups is heterogeneity among discoveries. The ideal experiment to assess when academic researchers commercialize their discoveries via startups would involve random matching of researchers and discoveries, which is of course impractical. Instead, we take advantage of the fact that different researcher teams sometimes make the same or very similar discoveries-which we label "twins".

3.1 Accounting for latent commercializability via "twin" discoveries

Bikard & Marx (2018) analyze 316 twin discoveries drawn from articles in the top 15 scientific journals during 2000-2010, which are primarily from the life sciences. We apply a similar technique to *all* published papers in the Web of Science from 1955-2017. We begin by finding all pairs of papers that a) were published no more than a year apart, b) are cited at least five times, c) share 50% of forward citations, and d) are both cited by at least one other paper. Applying these criteria to the Web of Science yields a list of 40,392 papers from 20,196 potential twin discoveries. The next step is to determine whether the potential twin discoveries are cited *adjacently* (i.e., within the same parenthesis). Adjacent citations suggest that forward-citing researchers are unable to attribute the discovery to a single paper, with the listed references within the citation parentheses receiving co-attribution.

Identifying adjacent citations involves inspecting the text of more than 1.2M papers that jointly cite what may be twin discoveries (although reference lists are available electronically, adjacent citation listings are not). Retrieving all such papers is impractical, as many if not most published articles reside behind pay-walls and are inaccessible at scale. However, PDFs of many papers are freely available—sometimes in draft form—and have been indexed by Google Scholar (GS). Although GS does not support bulk downloads, over a period of 19 months we retrieved approximately 280,000 publicly-available, non-paywalled PDFs corresponding to the 1.2M papers that jointly cited our 40,392 potential twin discoveries. For 29,257 of the 40,392 potential twin discoveries, we were able to determine whether they were adjacently cited by the PDFs that cited both of them. Of those, we found that 23,853 papers were cited adjacently. These comprise our population of twin discoveries, which should have similar latent commercializability among twins. Appendix A provides more detail on the twin discoveries, which hail from more than 3,000 academic institutions in 106 countries and span more than 200 scientific fields.

3.2 Outcome variable: entrepreneurial commercialization of scientific discoveries

Our dependent variable indicates whether academic researchers commercialize their scientific discoveries via a startup. To our knowledge, a large sample of academic scientific discoveries commercialized via startups has not been previously assembled. To be sure, many studies of technology transfer have tracked out-licensing or other forms of commercializing discoveries (Friedman & Silberman, 2003). Zucker, Darby, & Brewer (1998) examine the correlation between the presence of prominent scientists and entrepreneurial activity at the state level, but no direct connection between the scientist and startup is measured. Studies of academics-turned-entrepreneurs (Stuart & Ding, 2006) note when an academic either founded or advised a life-sciences startup but do not directly trace that involvement back to a particular discovery. In any case, prior studies have not considered the entire team of scientists involved with a particular discovery.

We measure entrepreneurial commercialization in two ways. First, we detect entrepreneurial commercialization via the U.S. Small Business Innovation Research (SBIR) grants. The SBIR program is targeted at encouraging "domestic small businesses to engage in federal research and research & development that has the potential for commercialization" and has awarded non-dilutive funding in excess of $45B since the program was initiated in 1982 (www.sbir.gov/about). We interpret pursuing SBIR funds as an indicator of commercialization aspirations. We calculate the pairwise overlap between scientists on a focal article and either the primary contact or principal investigator of SBIR awards two years before the publication of the article and up until five years thereafter. Scientists and SBIR personnel are compared individually, with an overall match score computed according to a) whether the surname is an exact vs. fuzzy match b) frequency of the surname in the Web of Science c) whether the middle initial matches (more details are provided in Appendix A). A weighted average of author/awardee overlap is computed to yield an overall article/SBIR match score. If multiple SBIR awards have identical author-overlap scores, we break ties with temporal proximity.

Our second method of determining entrepreneurial commercialization involves finding patent-paper pairs (Murray, 2002) where the patent is assigned to an entrepreneurial venture. The premise is that while scientific publications are the typical currency of academia, patents and their associated legal protection are valued much more in the commercial domain, and specifically by venture capitalists (Hsu & Ziedonis, 2013). Our algorithmic effort is therefore aimed at identifying patents which are granted to entrepreneurial ventures which cover the same or similar scientific advance in which there is overlap in authors. We start by finding the subset of twin academic discoveries that are cited by patents and check for overlap between the authors of the paper and the inventors named on the patent, following a process similar to that with papers and SBIR awardees. In some cases, the authors of an article have an identical overlap score with more than one patent. Ties are broken in two steps. First, the patent-paper pair closest in time (i.e., publication year vs. patent application year) is retained as in Thompson, Ziedonis, & Mowery, (2018). If two patents in the same year form pairs with the same paper, we further resolve ambiguity following (Magerman, et al. 2015) by choosing the patent-paper pair with the highest cosine similarity between the abstract of the article and the summary text of the patent. (Cosine similarity is computed using Term Frequency * Inverse Document Frequency, where all Web of Science abstracts and patent summaries are used as the corpus.)

However, not every patent-paper pair represents entrepreneurial commercialization. The scientists may merely patent the discovery but assign it to the university. Alternatively, one or more scientists on a paper may cooperate with an established firm to commercialize the discovery. We thus subset the list of patent-paper-pairs to those that are assigned to entrepreneurial ventures, as determined from VentureSource and CrunchBase. We find 139 academic articles that were commercialized via patent-paper pairs assigned to startups and 89 that were commercialized via SBIR awards, for a total of 228 entrepreneurial commercialization events. More details regarding construction of the outcome variable are in Appendix B. Appendix Table B1 also provides validation of the measure, confirming in a stratified random sample that both both patent-paper-pairs and overlapping SBIR grants truly reflect instances of a startup commercializing an academic discovery with the involvement of one of the original scientists.

3.3 Covariates

Given that our measure of entrepreneurial commercialization depends on an algorithm that scores the number of name matches (weighted for quality), twin discoveries with more authors might mechanically have higher overlap scores. Thus we control for the number of scientists on each article corresponding to the twin discovery. In addition to controlling for the number of scientists, we also enter covariates for the nature of the scientific team. First, we include the combined number of years the scientists on a focal paper have been research-active.

Second, we include a measure of the interdisciplinarity of the scientific team. This measure is calculated as one minus the Herfindahl-Hirsch index of the subjects for articles written by scientists on the focal paper. If all articles by all scientists on the focal article published all of their papers in the same subject, this variable would be set to zero.

Third, we measure whether the previous collaborators of the scientists on the paper include a "star" serial entrepreneurial commercializer. This variable is reminiscent of Stuart & Ding's (2016) measure of the number of prior collaborators who served as founders or advisory-board members of startups that filed for a IPO. Our measure differs in several ways. First, we measure involvement with early-stage ventures and not just those that complete an IPO. Second, instead of summing all instances of entrepreneurial involvement, we focus on "star" serial entrepreneurs who lie at the 99^{th} percentile of entrepreneurially-commercializing academic scientists in the year of the scientist's most recent collaboration. Third, instead of focusing on individual scientists we check whether *any* scientist on the paper had previously collaborated with such a star. Additional characteristics of 'star' commercializers are available in Appendix C.

Although the twin discoveries should have similar latent commercial potential, as found by Bikard & Marx (2018), individual articles may report the same discovery in more clinical or industry-relevant ways. We control for such factors by including the count of forward citations (in the next five years) to the focal article from patents assigned to corporations, as these may indicate that a particular article reporting the scientific discovery appear to have more commercial value than its twin. Patent-to-paper citations are computed following Fleming et al. (2018).

As noted earlier, entrepreneurial activity may depend on geography. We include a lagged count of venture investments in the same postal code as the focal discovery. We also include a lagged count of SBIR awards in the same postal code, which is by definition zero outside the U.S. Organizational characteristics may also influence commercializability. We control for the corresponding author's institutional research productivity in the same field as the paper.[3]

3.4 Descriptive statistics and estimation

Table 1 contains descriptive statistics and correlations. Table 2 shows difference-of-means tests between twin discoveries that were entrepreneurially commercialized vs. not. Discoveries that were commercialized by startups have more scientists on the team and more experienced scientists, and more interdisciplinary scientific teams. Commercialized discoveries have about 40% more citations from industry patents and, interestingly, are from institutions with somewhat fewer publications in the subject of the focal paper. Perhaps the most dramatic univariate difference is in the share of discoveries that have a 'star' commercializer among the scientists' prior collaborators. Barely one percent of non-commercialized discoveries have ties to such a star, compared with more than 10% of commercialized discoveries. Commercialized discoveries are moreover located in postal codes with more venture capital investments and SBIR awards.

[Tables 1 and 2 about here]

Following practice in epidemiological twin studies (Carlin et al., 2005), we estimate the likelihood of entrepreneurial commercialization using fixed effects for papers that report a twin discovery. The regression equation is:

$$ENTCOMM_{ij} = \alpha_0 + \alpha_1 STAR_i + \alpha_2 IDC_i + \alpha_3 X + \gamma_j + \varepsilon_{ij}$$

where *j* represents the twin discovery and *i* represents a paper reporting the twin discovery. $ENTCOMM_{ij}$ captures whether the focal article was commercialized by a startup. $STAR_i$ captures whether the scientists on a given article had previously collaborated with a "star" entrepreneurial commercializer. IDC_i reports the interdisciplinarity of the scientific team. X_i is a vector of other covariates. Finally, γ_j is a fixed effect for the twin discovery.

Our primary estimation approach utilizes linear probability models (LPM). Following Beck (2015), we also estimate conditional logit models, which exclude any duplicate discovery whether neither (or both)

[3] For institutions located in North America, we also have technology-transfer related variables from the Association of University Technology Managers and compute models limited to institutions where such variables are available. However, the AUTM data have substantial limitations, including a lack of data for the entire University of California system, so we do not report these models.

of the twin discoveries is commercialized. In robustness checks (Appendix D), we also estimate LPM models restricted to those twins where one article was commercialized and the other was not.

4. Results

We begin in Table 3 by evaluating the relationship between entrepreneurial commercialization and various groups of covariates. In column (1), it is evident that papers with larger teams of scientists are more likely to be commercialized via startups. This is expected, as our measure relies on name overlap between scientists and either patent inventors or SBIR recipients. Column (2) shows only an imprecise relationship between the institution's research productivity in the same field as the paper. Column (3) fails to precisely estimate the relationship between citations from patents assigned to corporations, which is reassuring as one might be concerned that our dependent variable could be conflated with articles simply being cited more often by patents.

[Table 3 about here]

Columns (4) and (5) add the key explanatory variables. Column (4) shows that discoveries where the scientific team is more interdisciplinary are more likely to be commercialized by startups. Column (5) shows that discoveries where the scientists' prior collaborators include a "star" commercializer are more likely to be commercialized via startups. Finally, column (6) examines the relationship between entrepreneurial commercialization and two geographic measures of the commercialization environment. We do not see a relationship between our dependent variable and either the previous-year count of venture-capital investments or the count of SBIR awards in the same postal code.

All of the foregoing covariates are included in column (7), which maintains statistical significance on the interdisciplinarity of the scientists as well as past collaboration with a 'star' commercializer. Using estimated coefficients from column (7) to calculate magnitudes, a one-standard-deviation increase in authorship team interdisciplinarity (0.27) corresponds to a 2.2% increase in the likelihood that a duplicate discovery will be commercialized by a startup. The presence of a "star" commercializer among the scientists' past collaborators is associated with a 7.5% increase. Robustness tests, including conditional logit estimation, are available in Appendix D.

In Table 4, we dig deeper into the nature of interdisciplinarity and 'stars.' Column (1) repeats column (7) of Table 4 to facilitate comparison. In columns (2-4) we explore how interdisciplinarity is involved with the entrepreneurial commercialization of science. Our primary measure captures the overall interdisciplinarity of the work conducted by the scientists on a focal paper, but this could be driven by several subfactors. In column (2), we replace the interdisciplinarity variable with a simple count of the primary disciplines represented by the scientists on the paper. (By "primary" discipline we mean the discipline in which each author publishes most often.) The positive, statistically-significant estimate of the associated coefficient suggests that having scientists from a variety of disciplines is important, not just having a set of scientists from the same discipline who also work relatively often in other areas.

[Table 4 about here]

That said, it does not appear crucial—or even advantageous—in the commercialization process for scientists to fully specialize. The covariate in column (3) of Table 4 counts the number of scientists who publish exclusively in a single field. If specialists were critical to the commercialization process, we might expect this coefficient to be precisely estimated, but it is not. Nor is it the case that it suffices to have one highly interdisciplinary scientist collaborating with a set of relative specialists. In column (4), we calculate each scientist's individual level of interdisciplinarity and then enter as a covariate the difference between the most interdisciplinary scientist and the mean of the team. The negative coefficient suggests that such a configuration does not facilitate commercialization. In other words, a set of relative specialists relying on a single boundary-spanner are less likely to commercialize than a set of scientists in a variety of disciplines who themselves are not overly specialized.

The remaining columns of Table 4 verify that it is the presence of a "star" *commercializer* among the scientists' past collaborators that explains the patterns in Table 3 and not simply an association with a highly prolific or highly-cited researcher. We test this alternative hypothesis in two ways. In column (5), we replace the 'star' commercializer variable—again, being in the 99^{th} percentile—with an indicator for having a past collaborator whose count of publications was in the 99^{th} percentile in the year of that most recent collaboration. Column (6) repeats this exercise with an indicator for whether any of those past collaborators was in the 99^{th} percentile of citations per article (in a five-year window following publication). Neither of these coefficients is estimated with much precision. We conclude that not just a star researcher but a star *commercializer* is necessary to facilitate entrepreneurial commercialization of science.

5. Discussion

By building a method to measure academic entrepreneurial commercialization attempts at scale and embedding the analysis in an empirical design to hold constant a given scientific advance, we investigate two team-based correlates of scientific-commercial translation. We find evidence that a team of more interdisciplinary scientists, as well as past collaboration with 'star' commercializers, predict the commercialization of academic scientific discoveries via startups.

A limitation of our methodology is that we may not capture scientific commercialization by a startup that licenses or otherwise appropriates the discovery without involvement from the original scientists. Given the prior literature about the importance of engaging the inventor and aligning incentives for commercialization success (e.g., Jensen & Thursby, 2001), we hope to spur more research in the academic startup channel of technology commercialization.

From a team design perspective, our results suggest that well-rounded individuals who are part of teams who are themselves well-rounded are more likely to pursue startup commercialization (a specific form of building team interdisciplinarity). While it remains to be seen whether this a general phenomenon beyond academia and the possible mechanisms underpinning the relationship, we hope future research explores this and related issues in greater detail.

Similarly, prior commercialization star affiliation holds a number of implications for attracting and retaining exceptional commercializers in organizations and institutions, and suggests a specific form of spillover to such relationships.

References

P. Azoulay, W. Ding, T. Stuart (2007). "The determinants of faculty patenting behavior: Demographics or opportunities?" *Journal of Economic Behavior & Organization*, 63: 599-623.

N. Beck (2015). "Estimating grouped data models with a binary dependent variable and fixed effects: what are the issues?" *Annual Meeting of the Society for Political Methodology.*

M. Bikard, M. Marx (2018). "Hubs as lampposts: academic location and firms' attention to science."

Carlin, John B., et al. "Regression models for twin studies: a critical review." *International Journal of Epidemiology*, 34.5 (2005): 1089-1099.

L. Fleming, G. Li, H. Greene, M. Marx, and D. Yao (2018). "U.S. innovation depends increasingly upon federal support".

J. Friedman, J. Silberman (2003). "University technology transfer: Do incentives, management, and location matter?" *Journal of Technology Transfer*, 28: 17-30.

J. Haltiwanger, RS. Jarmin, J. Miranda (2013). "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 95(2): 347-361.

L. Hong, SE. Page (2004). "Groups of diverse problem solvers can outperform groups of high-ability problem solvers," *Proceedings of the National Academy of Sciences*, 101(46): 16385-16389.

DH. Hsu, RH. Ziedonis (2013). "Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents," *Strategic Management Journal*, 34(7): 761-781.

R. Jensen, M. Thursby (2001). "Proofs and prototypes for sale: the licensing of university inventions," *American Economic Review*, 91(1): 240-259.

P. Krass (2018). "Edison of our times. Robert Langer share his thoughts on entrepreneurship," *Fortune*, August 1.

E. Leahey, CM. Beckman, TL. Stanko (2017). "Prominent but less productive: The impact of interdisciplinarity on scientists' research." *Administrative Science Quarterly*, 62(1), 105-139.

T. Magerman, B. Van Looy, K. Debackere (2015). "Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology," *Research Policy*, 44(9): 1702-1713.

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

R. Nanda, JS. Sorensen (2010). "Workplace peers and entrepreneurship," *Management Science*, 56(7): 1116-1126.

S. Samila, O. Sorenson (2011). "Venture capital, entrepreneurship, and economic growth," *Review of Economics and Statistics*, 93(1): 338-349.

PR. Sanberg, M. Gharib, PT. Harker, EW Kaler, RB Marchase, TD Sands, N. Arshadi, S. Sarkar (2014). "Changing the academic culture: valuing patents and commercialization toward tenure and career advancement," *Proceedings of the National Academy of Science*, 111(18): 6542-6547.

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

S. Shane (2001). "Technological opportunities and new firm creation," *Management Science*, 47(2): 205-220.

T. Stuart, W. Ding (2006). “When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences,” *American Journal of Sociology*, 112: 97-144.

T. Stuart, O. Sorenson (2003). “Liquidity events and the geographic distribution of entrepreneurial activity,” *Administrative Science Quarterly*, 48(2): 175-201.

NC. Thompson, AA. Ziedonis, DC. Mowery (2018). “University licensing and the flow of scientific knowledge,” *Research Policy*, 47(6): 1060-1069.

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

L. Zucker, M. Darby, M. Brewer (1998). “Intellectual human capital and the birth of US biotechnology enterprises,” *American Economic Review*, 88(1): 290-305.

Table 1: Descriptive statistics and correlations for 23,851 twin discoveries

	mean	stdev	min	max	1	2	3	4	5	6	7	8
1 # authors	6.23	4.65	1	30	1.000							
2 Ln paper authors' cumulative experience	1.64	1.29	0	6.77	0.501	1.000						
3 Ln institution # papers on topic	1.06	0.51	0	4.46	0.099	0.478	1.000					
4 Ln 5-yr citations from industry patents	0.08	0.32	0	4.37	0.051	-0.003	-0.046	1.000				
5 Interdisciplinarity of scientists' output	0.46	0.27	0	0.97	0.257	0.525	0.220	0.017	1.000			
6 Scientists' prior coauthors include 'star' commercializer	0.01	0.1	0	1	0.098	0.102	0.060	0.016	0.074	1.000		
7 Ln same-postalcode # investments (CB)	0.15	0.58	0	6.77	0.025	0.032	-0.039	0.065	-0.006	0.040	1.000	
8 Annual ln SBIR awards in postalcode	0.12	0.36	0	3.04	-0.029	-0.022	0.005	0.048	0.019	0.003	0.301	1.000

Table 2: Difference of means tests for duplicate discoveries that were commercialized by startups (n=224 of 23,851)

	commercialized	not commercialized	stderr	p<
# authors	8.991	6.207	0.311	0.000
Ln paper authors' cumulative experience	1.921	1.642	0.086	0.001
Ln institution # papers on topic	0.954	1.061	0.034	0.002
Ln 5-yr citations from industry patents	0.416	0.075	0.022	0.000
Interdisciplinarity of scientists' output	0.521	0.462	0.018	0.001
Scientists' prior coauthors include 'star' commercializer	0.112	0.009	0.006	0.000
Ln same-postalcode # investments (CB)	0.465	0.150	0.039	0.000
Annual ln SBIR awards in postalcode	0.187	0.123	0.024	0.008

Table 3 OLS estimates for startup-commercialization of 23,851 twin discoveries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# authors	0.000888**						0.000743*
	(0.000413)						(0.000412)
Ln paper authors' cumulative experience	0.000319						-0.000199
	(0.000843)						(0.00103)
Ln institution # papers on topic		-0.00128					-0.00396*
		(0.00179)					(0.00208)
Ln 5-yr citations from industry patents			0.00631				0.00696
			(0.0112)				(0.0111)
Interdisciplinarity of scientists' output				0.00967**			0.00839**
				(0.00398)			(0.00424)
Scientists' prior coauthors include 'star' commercializer					0.0756***		0.0746***
					(0.0242)		(0.0242)
Ln same-postalcode # investments (CB)						0.000767	0.000418
						(0.00288)	(0.00285)
Annual ln SBIR awards in postalcode						0.00423	0.00473
						(0.00289)	(0.00290)
Constant	0.00333	0.0108***	0.00890***	0.00491**	0.00867***	0.00875***	0.00350
	(0.00280)	(0.00199)	(0.00107)	(0.00194)	(0.000659)	(0.000759)	(0.00352)
R-squared	0.509	0.509	0.509	0.509	0.512	0.509	0.513

Note: fixed effects for each duplicate discovery and robust standard errors: *=p<.1; **=p<.05; ***=p<.01.

Table 4: Deeper examination of interdisciplinarity and prior collaboration with 'star' commercializers

	(1)	(2)	(3)	(4)	(5)	(6)
# authors	0.000703*	0.000461	0.000576	0.000518	0.000881**	0.000838**
	(0.000408)	(0.000405)	(0.000777)	(0.000426)	(0.000414)	(0.000412)
Ln paper authors' cumulative experience	-0.000295	-0.000334	0.000671	7.76e-05	0.000411	-0.000144
	(0.00102)	(0.00100)	(0.00107)	(0.00100)	(0.00105)	(0.00111)
Ln institution # papers on topic	-0.00419**	-0.00294	-0.00394*	-0.00418**	-0.00290	-0.00337
	(0.00208)	(0.00211)	(0.00207)	(0.00208)	(0.00209)	(0.00208)
Ln 5-yr citations from industry patents	0.00735	0.00735	0.00729	0.00738	0.00611	0.00617
	(0.0110)	(0.0110)	(0.0110)	(0.0110)	(0.0112)	(0.0112)
Ln same-postalcode # investments (CB)	0.000422	0.000403	0.000419	0.000444	0.000740	0.000732
	(0.00284)	(0.00285)	(0.00285)	(0.00284)	(0.00288)	(0.00288)
Annual ln SBIR awards in postalcode	0.00474	0.00485*	0.00479*	0.00477*	0.00450	0.00447
	(0.00289)	(0.00290)	(0.00289)	(0.00289)	(0.00290)	(0.00290)
Interdisciplinarity of scientists' output	0.00839**				0.00895**	0.00854**
	(0.00423)				(0.00426)	(0.00425)
# scientists' primary disciplines represented		0.00368**				
		(0.00151)				
# scientists who publish only in one subject			0.000173			
			(0.000844)			
Difference in max & mean scientist interdisciplinarity				-0.0123*		
				(0.00659)		
Scientists' prior collaborators includes 'star' commercializer	0.0930***	0.0928***	0.0932***	0.0934***		
	(0.0256)	(0.0256)	(0.0256)	(0.0256)		
Scientists' prior collaborators includes one in 99th percentile productivity					-0.00285	
					(0.00249)	
Scientists' prior collaborators includes one in 99th percentile of cites per article						0.00107
						(0.00231)
Constant	0.00394	0.00170	0.00608*	0.0168**	0.00178	0.00247
	(0.00348)	(0.00396)	(0.00328)	(0.00694)	(0.00356)	(0.00353)
R-squared	0.514	0.514	0.514	0.514	0.510	0.510

Note: All models estimated w/OLS; fixed effects for each duplicate discovery; robust standard errors: *=p<.1; **=p<.05; ***=p<.01.

Appendix A: Characteristics of Twin Discoveries

Our 23,851 twin discoveries range from 1973-2015 and are from more than 3,000 academic institutions in 106 countries. Figure A1 shows their temporal distribution. (There may be additional twin discoveries in the distant past, but these are hard to discover because SBIR data are available only since 1983, and patent-to-paper citations are difficult to collect pre-1976 given errors in OCR processing of patent applications.)

Figure A1: Temporal Distribution of Twin Discoveries

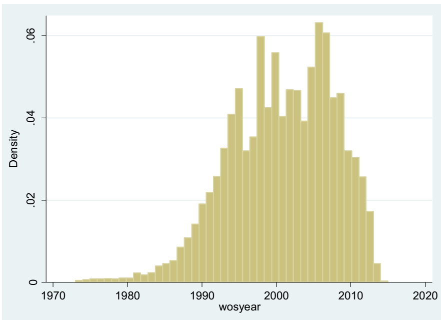

Table A1 shows the distribution of twin discoveries by geography, discipline, and institution. Over half of twin discoveries occur in the U.S., followed by Great Britain, Germany, and Japan. When considering pairs of twin papers, one-third of pairs both occur in the U.S. and 37% of twin papers are in the same country.

Panel B details the disciplinary fields of the twin discoveries. The life sciences are responsible for many of the most popular categories of twin discoveries, although Physics is the most popular category. Astronomy & Astrophysics is also a frequent source of twin discoveries. Finally, Panel C tabulates the academic institutions with the most twin discoveries.

Table A1: Twin Geography, Disciplines, and Institutions

Panel A		Panel B		Panel C	
Top 20 countries	%	Top 20 disciplines	%	Top 20 institutions	%
United States	54.1	Physics	6.0	Harvard	3.3
Great Britain	8.1	Cell Biology	5.4	UC San Francisco	1.5
Germany	6.8	Medicine, General & Internal	4.8	Stanford	1.5
Japan	5.3	Genetics & Heredity	4.0	University of Texas	1.4
France	4.5	Immunology	3.7	MIT	1.3
Canada	3.2	Astronomy & Astrophysics	2.9	UC Berkeley	1.3
Netherlands	2.1	Neurosciences	2.9	Yale	1.3
Italy	2.1	Oncology	2.6	Johns Hopkins	1.1
Switzerland	2.0	Developmental Biology	2.0	UC San Diego	1.1
Austria	1.7	Hematology	1.6	Caltech	1.0
Sweden	1.2	Physics, Condensed Matter	1.5	Columbia	0.9
China	1.1	Cardiac & Cardiovascular Systems	1.5	UCLA	0.9
Israel	0.9	Clinical Neurology	1.3	Cambridge University	0.9
Spain	0.7	Chemistry	1.2	Washington University	0.9
Denmark	0.7	Virology	1.1	University of Washington	0.9
Austria	0.6	Endocrinology & Metabolism	1.0	Tokyo University	0.9
Belgium	0.4	Geochemistry & Geophysics	1.0	University of Pennsylvania	0.8
Finland	0.4	Gastroenterology & Hepatology	0.9	University of Michigan	0.8
South Korea	0.4	Optics	0.9	Oxford University	0.8
Scotland	0.2	Chemistry, Physical	0.8	Rockefeller University	0.8

Appendix B: Details of constructing the startup commercialization outcome variable

As we state in the main text, our outcome variable of startup commercialization is measured in two ways: (a) scientist involvement in the US SBIR program, and (b) patent-paper pairs in which a patent is assigned to an entrepreneurial venture (which also cites the focal research paper and contains author overlap across the patents and papers. The purpose of this appendix is to provide more details about variable construction and to report result robustness.

The U.S. SBIR program is a program which requires U.S. federal agencies (which have research expenditures in excess of $100M) to set aside currently 3.2% of their budget (as of fiscal year 2017; this rate has varied over time) to award non-dilutive funding to U.S. small businesses which meet its mission and goals (as articulated in the main text). The US federal agencies participating in the program include: Department of Agriculture, Department of Commerce; Department of Defense; Department of Education; Department of Energy; Department of Health and Human Services; Department of Homeland Security; Department of Transportation; Environmental Protection Agency; National Aeronautics and Space Administration; and National Science Foundation.

While each agency administers its own individual program, awards are made on a competitive basis following proposal evaluation. There are three phases to the SBIR program: Phase I is to "establish technical merit, feasibility, and commercial potential…" and such awards "normally do not exceed $150,000 total costs for 6 months." Phase II awards are to: "…continue the R/R&D efforts initiated in Phase I…Only Phase I awardees are eligible for a Phase II award…[and] normally do not exceed $1M total costs for 2 years." Phase III is for small businesses to pursue commercialization objectives, though the SBIR program does not fund Phase III development.

We implement name matching for Web of Science authors vs. SBIR personnel, removing hyphenation and other punctuation. (We examine the first 30 authors on each paper although some papers have more than 30 authors.) Although full names are available for SBIR and patents, many papers only have the authors' surname and initial(s). If both the author and the SBIR awardee have both initials present but these do not match, a score of zero is assigned. Names lacking first initials are ignored. Otherwise, a match score is assigned through a series of steps. First, we determine whether the surnames match exactly or nearly, where "nearly" indicates that both surnames are more than five characters long and fewer than ¼ of the characters must be changed to convert one to the other (i.e., Levenshtein distance). Moreover, the surnames must start with the same letter (e.g., "Rogers" and "Bogers" are not matched). Two names are treated as a preliminary match if the surname meets these criteria and the first initials also match. We want to avoid the situation where the author "J Smith" is assumed to the be same as the SBIR awardee "Jesse Smith", so we score surnames according to their inverse frequency of appearance in the Web of Science. For instance, surname Smith would be downscaled to near-zero as it is among the most common author names. Surnames that comprise less than 0.007% of all authors (i.e., 2nd percentile) are not downscaled. If only two authors match between the paper and SBIR grant, and both of them represent more than 0.005% of all authors, we conclude that there is no match. Regardless of surname, matches are considered exact if both first and second initials are present for both names and they both match. A similar algorithm is implemented for computing overlap between authors of articles and inventors on patents.

Finally, in identifying unique authors we initially relied on the Web of Science author ID. However, in our testing we found that many scientists with different surnames were grouped under a single ID. We split author IDs based on different surnames or, if surnames matched, different first and (if available) middle initials. Doing so raised the number of authors in the Web of Science from approximately 1 million to about 73 million.

To evaluate whether our algorithm truly captures instances of startup commercialization, we examine a small random sample of both types of potential examples of commercialization to seek direct confirmation of our algorithmic approach. Panel A of Table B1 shows the examples of paper-patent pairs, and Panel B shows the examples of SBIR grants. We start by randomly selecting five scientific papers drawn from each route of identifying commercialization. For each of these papers, we retrieve the underlying scientific article via Google Scholar searches and record the authors. For Panel A, we retrieve the associated patent from our algorithmic approach described in the main text via Google Patents (patents.google.com). We record the patent title, inventors, and assignee. For Panel B, we retrieve the associated SBIR grants to the focal companies via sbir.gov and record the grant title, funding agency and amount, and the listed principal investigator/business contact. To verify the linkages in both panels between scientific paper and commercialization activity, we conduct web searches in the following manner: we find the overlapping names between paper author and patent inventor (Panel A) or SBIR contact (Panel B) – those are shown in bold in the table. We search the web for the union of the overlapped name(s) and the new venture entity (patent assignee in Panel A; SBIR company in Panel B). The final column in both panels of the table provide web links (all accessed in January 2019) providing confirmation of commercialization activity in all ten instances.

One interesting case is the second entry in Panel A. We initially had difficulty finding confirmation, but then found that one of the author/inventors, Larry Gold, had founded a company, NeXagen to commercialize his technology, changed the name of the company, and subsequently sold that company to Gilead Sciences. The patent was subsequently reassigned to Gilead Sciences, which is why initially we thought we had failed to find a linkage.

[Appendix Table B1 about here]

Appendix Table B1, Panel A: random sample of five patent-paper-pair instances of startup commercialization

Paper title	Journal / Year	Authors	Institution	Patent	Inventors	Patent assignee	Linkages
RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions	*PNAS* / 2011	Wiedenheft, B; van Duijin, E; Bultema, JB; Waghmare, SP; Dickman, M; Zhou, KH; Barendregt, A; Westphal, W; **Doudna, JA**	Univ Calif Berkeley	Compositions and methods of nucleic acid-targeting nucleic acids (9260752	Andrew Paul May; **Rachel E. Haurwitz**; **Jennifer A. Doudna**; James M. Berger; Matthew Merrill Carter; Paul Donohoue	Caribou Biosciences, Inc.	**Doudna** is on Caribou's SAB; **Haurwitz** is Caribou's CEO and on the firm's BoD. Source: https://cariboubio.com/about-us
Systematic evolution of ligands by exponential enrichment - RNA ligans to bacteriophage-T4 DNA-polymerase	*Science* / 1990	Tuerk, C; **Gold, L**	Univ Colorado	Systematic evolution of ligands by exponential enrichment: tissue selex (6613526)	Joseph S. Heilig; **Larry Gold**	Gilead Sciences, Inc.	**Gold** is a founder of NeXagen, which became NeXstar Pharmaceuticas. That organization merged with Gilead Sciences in 1999. Source: https://somalogic.com/about-us/leadership/larry-gold-2/
Phase selection of microcrystalline GaN synthesized in supercritical ammonia	*Journal of Crystal Growth* / 2006	**Hashimoto, T**; Fujito, K; Sharma, R; **Letts, ER**; Fini, PT; Speck, JS; Nakamura, S	Univ Calif Santa Barbara	Method for producing group III-nitride wafers and group III-nitride wafers (9803293)	**Tadao Hashimoto**; Edward Letts; Masanori Ikari	SixPoint Materials Inc	**Hashimoto** is CEO/CTO of SixPoint; **Letts** is VP of Technology of the firm. Source: http://www.spmaterials.com/team.htm
Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology	*NEJM* / 2012	Alexander, EK; **Kennedy, GC;** Baloch, ZW; Cibas, ES; Friedman, L; Lanman, RB; Mandel, SJ ;Yener, N; Kloos, RT; LiVolsi, VA; Lanman, RB; Steward, DL; Friedman, L; Kloos, RT; Wilde, JI; Raab, SS; Haugen, BR; Steward, DL; Zeiger, MA; Haugen, BR	Brigham & Womens Hospital	Algorithms for disease diagnostics (9495515)	**Giulia C. Kennedy**; Darya I. Chudova; Eric T. Wang; Jonathan I. Wilde	Veracyte Inc	**Kennedy** is Chief Scientific and Medical Officer of Veracyte. https://www.veracyte.com/who-we-are/leadership/executive-team. Wilde was a director and VP of Discovery Research at Veracyte. https://uk.linkedin.com/in/jonathanwilde650
Human retinoblastoma susceptibility gene - cloning, identification, and sequence	*Science* / 1987	**Lee, WH**; Bookstein, R; Hong, F; Young, LJ; Shew, JY; Lee, EYHP	Univ Calif San Diego	Therapeutic use of the retinoblastoma susceptibility gene product (5851991)	**Wen-Hwa Lee**; Eva Y-H.P. Lee; David W. Goodrich; H. Michael Shepard; Nan Ping Wang; Duane Johnson	University of California; Canji Inc	**Wen-Hwa Lee** was Chair of the Scientific Advisory Board of Canji, Inc. http://rcndd.cmu.edu.tw/sites/default/files/WHL-CV.pdf. Canji was "formed to commercialize suppressor oncogene technology developed by Dr. Wen-Hwa Lee of the University of California at San Diego. Canji, Inc. operates as a subsidiary of Merck & Co." https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapid=26032.

Appendix Table B1, Panel B: random sample of five SBIR instances of startup commercialization

Paper title	Journal / Year	Authors	Institution	SBIR Company	SBIR Grant(s)	SBIR PIs	Linkages
The outer mitochondrial membrane protein mitoNEET contains a novel redox-active 2Fe-2S cluster	*Journal of Biological Chemistry* / 2007	Wiley, SE; Paddock, ML; Abresch, EC; Gross, L; van der Geer, P; Nechushtai, R; **Murphy, AN**; Jennings, PA; Dixon, JE	Univ Calif San Diego	Mitokor, Inc.	"Mitochondrial Functional Proteomics" (2005 for $100,000 from the Department of Defense); "Osteoarthritis/Chondrocalcinosis: Mitochondrial Therapy" ($106,745 from the Department of Health and Human Services (HHS))	Eoin Fahy; **Anne Murphy**	**Murphy** was Director of Mitochondrial Biology at MitoKor: https://www.researchgate.net/profile/Anne_Murphy/2
Scattering theory derivation of a 3D acoustic cloaking shell	*Physical Review Letters* / 2008	Cummer, SA; Popa, B; Schurig, D; Smith DR; Pendry, J; Rahm, M; **Starr A**	Duke Univ	SensorMetrix, Inc.	"Development of Acoustic Metamaterial Applications" ($750,813 from the Dept of Defense (Navy))	**Anthony Starr**	Dr. **Anthony Starr** is the founder, president & CEO of SensorMetrix. http://www.sensormetrix.com/key-personnel.html
Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini	*Cell* / 2008	Mahrus, S; Trinidad, JC; Barkan, DT; Sali, A; Burlingame, AL; **Wells, JA**	Univ Calif San Francisco	Sunesis Pharmaceuticals, Inc.	"Development of Conformation Specific Kinase Inhibitors" (HHS for $1.5M)	**James A. Wells**	**Wells** is founder of Sunesis Pharmaceuticals. https://www.crunchbase.com/person/jim-wells#section-jobs and https://www.bloomberg.com/research/stocks/private/person.asp?personId=467474&privcapId=3768647&previousCapId=177932577&previousTitle=REZOLUTE%20INC
Curved plasma channel generation using ultraintense airy beams	*Science* / 2009	Polynkin, P; Kolesik, M; **Moloney, JV**; Siviloglou, GA; Christodoulides, DN	Univ Arizona	Nonlinear Control Strategies, Inc.	"High Power, Room Temperature 2.4- 4 micron Mid-IR Semiconductor Laser Optimization" (Department of Defense (Air Force) for $99,995 and $746,925	**Jerome V Moloney**	**Moloney** is President and corporate head of Nonlinear Control Strategies. http://www.nlcstr.com/contact.htm
Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma	*Journal of Experimental Medicine* / 2012	**Kiel, MJ**; Velusamy, T; Betz, BL; Zhao, L; Weigelin, HG; Chiang, MY; Huebner-Chan, DR; Bailey, NG; Medeiros, LJ; Bailey, NG; Elenitoba-Johnson, KSJ	Univ Michigan	Genomenon, Inc.	"Commercial Software Using High throughput Computational Techniques to Improve Genome Analysis" (HHS- National Institutes of Health, $972,083)	**Mark Kiel**	**Kiel** is a co-founder of Genomenon and Chief Science Officer. https://www.genomenon.com/about/; https://www.crunchbase.com/organization/genomenon

Appendix C: Characteristics of 'Star' Commercializers

Appendix Table C1 provides additional information on the nature of "star" entrepreneurial commercializers. Only 0.4% of the more than 73 million authors in the Web of Science have had one of their discoveries commercialized by a startup. The vast majority of authors whose discoveries are commercialized by startups do so only once (mean = 1.26). Among commercializing authors, 2.4% are among the top 1% of authors whose discoveries are commercialized by startups during at least one year of their career. Overall, less than 0.01% of all authors are ever "stars" in this respect.

Panel A of Appendix Table C1 compares stars with all other authors in the Web of Science. Perhaps unsurprisingly, stars have many more articles and citations per article, and they have been publishing longer than non-stars. Panel B details the most popular fields among stars, using 251 fields from the Web of Science. Biochemistry & Molecular Biology is the most frequent field for entrepreneurial commercialization (13.2% of all stars work primarily in this field), followed by Chemistry, Electrical & Electronic Engineering, Immunology, and Applied Physics.

Appendix Table C1: Descriptive statistics for "star" entrepreneurial commercializers

Panel A: Star commercializers vs. all other authors (n-7,164 vs. 73,923,279)

	avg. non-star	avg. star	stderr	p<
lifetime # articles	1.639	13.708	0.040	0.000
average citations per paper	13.179	30.961	0.555	0.000
# years publishing	0.899	7.423	0.035	0.000

Panel B: Most popular fields for "star" startup-commercializers (all above 1%)

Field of Study	% of stars
Biochemistry & Molecular Biology	13.2%
Chemistry, Multidisciplinary	6.5%
Engineering, Electrical & Electronic	5.1%
Immunology	4.5%
Physics, Applied	4.2%
Oncology	3.9%
Multidisciplinary Sciences	3.6%
Chemistry, Medicinal	3.6%
Cardiac & Cardiovascular Systems	3.3%
Endocrinology & Metabolism	2.9%
Biotechnology & Applied Microbiology	2.7%
Biochemical Research Methods	2.7%
Optics	2.3%
Hematology	2.2%
Pharmacology & Pharmacy	1.9%
Chemistry, Physical	1.7%
Gastroenterology & Hepatology	1.7%
Neurosciences	1.6%
Urology & Nephrology	1.5%
Clinical Neurology	1.3%
Engineering, Biomedical	1.3%
Genetics & Heredity	1.3%
Radiology, Nuclear Medicine & Medical Imaging	1.2%
Chemistry, Organic	1.1%
Ophthalmology	1.1%

Appendix D: Robustness tests

Appendix Table D1 contains robustness checks and placebo tests, with column (1) repeating column (7) of Table 3 for convenience. Column (2) re-estimates column (1) in a conditional logit framework. Because the maximum likelihood estimator drops any groups without variation in the dependent variable, the inclusion of fixed effects on each twin discovery renders the number of observations much smaller. Statistical significance is reduced somewhat for both the 'star' and interdisciplinary result (to the 2% and 7% levels, respectively). Following Beck (2015), in column (3) we compare logit and OLS specifications by limiting the observations in OLS to the set of duplicate discoveries with variation in the outcome variable (which the maximum likelihood estimator does automatically). Unsurprisingly, results closely resemble that of the logit estimates.

In column (4) of Appendix Table D1, we randomly generate values of the dependent variable, which yields no statistical significance on any covariates. In unreported results, this placebo test also fails if the distribution of the randomly-generated dependent variable matches that of the actual dependent variable (i.e., less than 1% of papers are commercialized by startups).

Appendix Table D1: Robustness tests

	(1)	(2)	(3)	(4)
DV =	commercialization via startup			randomly generated
# authors	0.000743*	0.0325	0.0144	-0.00128
	(0.000412)	(0.0264)	(0.0119)	(0.00150)
Ln paper authors' cumulative experience	-0.000199	-0.0790	-0.0358	-0.000938
	(0.00103)	(0.133)	(0.0614)	(0.00569)
Ln institution # papers on topic	-0.00396*	-0.436	-0.192	-0.0137
	(0.00208)	(0.279)	(0.119)	(0.0116)
Ln 5-yr citations from industry patents	0.00696	0.148	0.0633	-0.0275
	(0.0111)	(0.233)	(0.104)	(0.0235)
Interdisciplinarity of scientists' output	0.00839**	1.015*	0.468*	-0.0123
	(0.00424)	(0.565)	(0.260)	(0.0239)
Scientists' prior coauthors include 'star' commercializer	0.0746***	1.172**	0.487***	0.0138
	(0.0242)	(0.473)	(0.164)	(0.0446)
Ln same-postalcode # investments (CB)	0.000418	0.0534	0.0229	-0.00783
	(0.00285)	(0.113)	(0.0522)	(0.00854)
Annual ln SBIR awards in postalcode	0.00473	0.402	0.184	0.00502
	(0.00290)	(0.282)	(0.128)	(0.0136)
Constant	0.00350		0.298**	0.537***
	(0.00352)		(0.146)	(0.0149)
Observations	23,851	436	436	23,851
Model	OLS	cond. logit	OLS	OLS
Adjusted R-squared		0.0111		
R-squared	0.513		0.093	0.592

Note: Column (3) restricts estimation to twin discoveries where only one of the pair was startup-commercialized. All models have fixed effects for each duplicate discovery and robust standard errors: *=p<.1; **=p<.05; ***=p<.01